

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
MESTRADO EM CIÊNCIA FLORESTAL

EULÁLIA APARECIDA SILVA

**CLASSIFICAÇÃO DA CAPACIDADE PRODUTIVA DE POVOAMENTOS DE
EUCALIPTO POR MEIO DE MÉTODOS TRADICIONAIS E REDES KOHONEN**

DIAMANTINA
2017

EULÁLIA APARECIDA SILVA

**CLASSIFICAÇÃO DA CAPACIDADE PRODUTIVA DE POVOAMENTOS DE
EUCALIPTO POR MEIO DE MÉTODOS TRADICIONAIS E REDES KOHONEN**

Dissertação apresentada ao Programa de
Pós-Graduação em Ciência Florestal da
Universidade Federal dos Vales do
Jequitinhonha e Mucuri, como requisito
parcial para obtenção do título de Mestre.

Orientador: Dr. Marcio Leles Romarco de Oliveira

**DIAMANTINA
2017**

Ficha Catalográfica – Serviço de Bibliotecas/UFVJM
Bibliotecário Anderson César de Oliveira Silva, CRB6 – 2618.

S586c Silva, Eulália Aparecida
Classificação da capacidade produtiva de povoamentos de eucalipto
por meio de métodos tradicionais e redes Kohonen / Eulália Aparecida
Silva. – Diamantina, 2017.
81 p. : il.

Orientador: Marcio Leles Romarco de Oliveira

Dissertação (Mestrado – Programa de Pós-Graduação em Ciência
Florestal) - Universidade Federal dos Vales do Jequitinhonha e Mucuri.

1. Redes neurais artificiais. 2. Curva-guia. 3. Predição dos
parâmetros. 4. Equação das diferenças. I. Título. II. Universidade
Federal dos Vales do Jequitinhonha e Mucuri.

CDD 634.973

Elaborado com os dados fornecidos pelo(a) autor(a).

EULÁLIA APARECIDA SILVA

**Classificação da capacidade produtiva de povoamentos de eucalipto
por meio de métodos tradicionais e redes Kohonen**

Dissertação apresentada ao
PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIA FLORESTAL - STRICTO
SENSU, nível de MESTRADO como
parte dos requisitos para obtenção do
título de MAGISTER SCIENTIAE EM
CIÊNCIA FLORESTAL

Orientador : Prof. Dr. Marcio Leles
Romarco De Oliveira

Data da aprovação : 18/07/2017


Prof.Dr. HELIO GARCIA LEITE - UFV


Prof.Dr. ALESSANDRO VIVAS ANDRADE - UFVJM


Prof.Dr. MARCIO LELES ROMARCO DE OLIVEIRA - UFVJM

DIAMANTINA

Dedico

A Deus.

À minha família.

AGRADECIMENTOS

Agradeço a Deus pela minha vida, por ser a razão da minha existência e por ser “a destra que me sustenta”.

Agradeço à minha família, por ser o bem mais precioso que eu poderia ter. Obrigada pelo incentivo constante e por acreditarem nos meus projetos. Agradeço especialmente aos meus pais Maria Aparecida e José Alberto, que com toda simplicidade me ensinaram as lições mais importantes da vida: princípios, valores, fé e amor. Esse alicerce bem edificado norteará minhas ações em qualquer lugar, para onde quer que eu for.

Agradeço ao professor Marcio Leles que me orientou por quatro anos, muito obrigada! Marcio contribuiu muito para minha formação, não apenas no desenvolvimento dos trabalhos de pesquisa, mas também nas disciplinas por ele ministradas. Para mim, sempre será referência na dedicação ao trabalho e na busca por fazer o seu melhor no desempenho de suas atividades.

Agradeço aos professores Helio Garcia e Alessandro Vivas pela disponibilidade em participar da banca e por suas contribuições para o trabalho.

Agradeço aos meus amigos Lidia, Cristiano, John, Lílían, Késsia e Raquel. Foi muito importante para mim poder contar com vocês. É bom saber que conquistei mais que um título acadêmico no Mestrado. Que a nossa amizade resista ao tempo e à distância!

Agradeço ao pessoal do Laboratório de Mensuração e Manejo Florestal pelas contribuições, em especial ao José Lucas pela ajuda para determinar os parâmetros iniciais para ajuste do modelo logístico e pela disponibilidade em esclarecer dúvidas.

Agradeço aos colegas da pós-graduação pelo companheirismo.

Agradeço à CAPES pela concessão da bolsa que me permitiu concluir o mestrado.

Agradeço à UFVJM pela minha formação profissional.

A todos que contribuíram de alguma forma com este trabalho, muito obrigada!

RESUMO

O objetivo do trabalho foi avaliar a eficiência da classificação da capacidade produtiva de povoamentos florestais de eucalipto (*Eucalyptus* spp.) por meio de rede neural artificial (RNA). Os dados utilizados foram provenientes de inventários florestais contínuos conduzidos em povoamentos de clones de *Eucalyptus* spp. localizados no estado de Minas Gerais. A classificação da capacidade produtiva foi realizada por meio de quatro métodos: curva-guia, predição dos parâmetros, equação das diferenças e rede neural artificial. Em todos os métodos foi adotada uma idade de referência de 72 meses e foram obtidas três classes de capacidade produtiva (superior, média e inferior). Para os métodos da curva-guia e equação das diferenças foi empregado o modelo de Schumacher linearizado e para o método da predição dos parâmetros foi utilizado o modelo logístico. Na classificação por meio de RNA utilizou-se a rede auto-organizável de Kohonen, sendo o agrupamento realizado em dois estágios. Na primeira etapa os dados foram utilizados para treinar a rede e na segunda etapa os vetores de pesos sinápticos foram agrupados utilizando o método do vizinho mais distante. Foram testadas diferentes entradas (E) para as RNA: E1- volume total com casca (*V*); E2- área basal (*B*); E3- altura total (*Ht*); E4- altura dominante (*Hd*); E5- diâmetro quadrático médio (*q*); e E6- *V*, *B*, *Ht*, *Hd*, *q* e número de árvores por hectare. A seleção da entrada foi realizada por meio da análise discriminante, sendo selecionada a entrada E6 com 83,6% de acerto geral. Os métodos foram comparados em termos de porcentagem de coincidência na alocação dos talhões, área e volume por classe de capacidade produtiva. As classes obtidas pelos métodos da curva-guia e equação das diferenças foram muito semelhantes de acordo com os critérios de comparação adotados. A classificação pelo método da predição dos parâmetros não foi semelhante aos outros métodos. A classificação por meio de rede neural artificial foi eficiente quando comparada aos demais métodos em termos de porcentagem de coincidência na alocação dos talhões, área e estoque volumétrico por classe de capacidade produtiva.

Palavras-chave: Redes Neurais Artificiais. Curva-guia. Predição dos Parâmetros. Equação das Diferenças.

ABSTRACT

The objective of this work was to evaluate the efficiency of the classification of the productive capacity of eucalyptus forest stands (*Eucalyptus* ssp.) through artificial neural network (ANN). The data used came from continuous forest inventories conducted in stands of *Eucalyptus* ssp. located in the state of Minas Gerais. The classification of the productive capacity was accomplished through four methods: guide curve, prediction of the parameters, equation of the differences and artificial neural network. In all methods, a reference age of 72 months was adopted and three productive capacity classes (upper, middle and lower) were obtained. For the methods of the guide curve and equation of the differences was used the linearized Schumacher model and for the method of the prediction of the parameters was used the logistic model. In the classification by ANN was used the self-organizing network Kohonen, which the grouping was performed in two stages. In the first step the data were used to train the network and in the second step the vectors of synaptic weights were grouped using the method of the most distant neighbor. Different entries (E) for RNA were tested: E1- total volume with bark (V); E2- basal area (B); E3- total height (Ht); E4- dominant height (Hd); E5- square mean diameter (q); and E6- V , B , Ht , Hd , q and number of trees per hectare. The selection of the input was performed through the discriminant analysis, and the E6 input was selected with 83.6% of the general hit. The methods were compared in terms of coincidence percentage in the allocation of stands, area and volume by class of productive capacity. The classes obtained by the guide curve methods and equation of the differences were very similar according to the adopted criteria of comparison. Classification by the method of parameter prediction was not similar to the other methods. The classification by artificial neural network was efficient when compared to the other methods in terms of coincidence percentage in the allocation of stands, area and volumetric stock by productive capacity class.

Keywords: Artificial Neural Networks. Guide Curve. Prediction of Parameters. Equation of differences.

SUMÁRIO

1 INTRODUÇÃO.....	9
2 OBJETIVOS.....	12
2.1 Objetivo geral	12
2.2 Objetivos específicos.....	12
3 REVISÃO DE LITERATURA	13
3.1 Capacidade produtiva de sítios florestais	13
3.2 Por que classificar a capacidade produtiva.....	13
3.3 Métodos de classificação da capacidade produtiva	14
3.3.1 Classificação pelo índice de local.....	16
3.3.1.1 Altura dominante	17
3.3.1.2 Tipos de curvas de índice de local.....	18
3.3.1.3 Construção de curvas de índice de local.....	19
3.3.1.4 Estabilidade das curvas de índice de local.....	21
3.3.1.5 Testes de anamorfismo	21
3.4 Novas ferramentas para classificação da capacidade produtiva	22
3.5 Redes neurais artificiais.....	22
3.5.1 Mapas auto-organizáveis de kohonen.....	25
3.5.1.1 Treinamento sequencial do mapa auto-organizável	26
3.5.1.2 Visualização do mapa treinado.....	30
3.5.1.2.1 Matriz de distância unificada (matriz-U).....	30
3.5.1.2.2 Planos de componentes.....	31
3.5.1.2.3 Histograma	32
3.5.1.3 Medidas de erro	32
3.5.1.4 Segmentação do mapa treinado	33
4 MATERIAIS E MÉTODOS.....	36
4.1 Descrição dos dados	36
4.2 Classificação da capacidade produtiva	37
4.2.1 Métodos da curva-guia, predição dos parâmetros e equação das diferenças.....	38
4.2.1.1 Testes de anamorfismo	42
4.2.2 Rede neural artificial	42
4.3 Comparação dos métodos	45

4.4 Análises Estatísticas	45
5 RESULTADOS E DISCUSSÃO	46
5.1 Análise exploratória dos dados	47
5.1 Método da curva-guia	47
5.2 Método da predição dos parâmetros	52
5.3 Método da equação das diferenças	56
5.3 Teste de anamorfismo	58
5.4 Redes neurais artificiais	61
5.5 Comparação dos métodos	68
6 CONCLUSÕES	74
REFERÊNCIAS	75

1 INTRODUÇÃO

O manejo de florestas requer que uma série de condições sejam avaliadas e consideradas, do planejamento à implementação das prescrições (SKOVSGAARD e VANCLAY, 2008). Nesse processo, são avaliadas as consequências de diferentes ações de manejo. Para isso os gestores florestais sintetizam informações sobre a condição atual e prevista de uma floresta por meio de indicadores (SKOVSGAARD e VANCLAY, 2008). Um desses indicadores é capacidade produtiva das florestas, isto é, o potencial para produção de madeira ou outro produto (CAMPOS e LEITE, 2013).

A classificação das propriedades florestais quanto a seu potencial produtivo é, portanto, referência para diagnóstico e prescrição de manejo, servindo de base para o planejamento de curto e longo prazo (SCOLFORO, 1992). Dessa forma, a classificação da capacidade produtiva é essencial para o manejo de florestas e constitui um dos primeiros e mais importantes passos para o planejamento das atividades florestais (OLIVEIRA, 1998).

A capacidade produtiva de um local pode ser analisada por meio de fatores do ambiente que tenham estreita relação com o crescimento do povoamento florestal (método indireto), ou então mensurando diretamente esse crescimento como resultado das condições ambientais (método direto) (BRUM, 1979).

Dentre os métodos diretos de classificação da capacidade produtiva encontrados na literatura, os mais comuns são os que expressam a capacidade por meio do índice de local, que é a altura dominante média do povoamento em uma idade específica, denominada idade índice (CAMPOS e LEITE, 2013).

A altura dominante tem sido preferida como indicadora da qualidade do local, pois, dentro de certos limites, não é influenciada pela densidade do povoamento e intervenções silviculturais (GARCIA, 1995). Clutter et al. (1983) também destacam que para muitas espécies existe uma correlação positiva entre a altura dominante e a produção volumétrica. Além disso, essa variável apresenta melhor correlação com as condições ambientais e maior praticidade na aplicação (BRUM, 1979).

Uma desvantagem de classificar a capacidade produtiva por meio da altura dominante é que as medições de altura são frequentemente inexatas. Isso ocorre porque a altura é medida indiretamente e em muitas situações a visibilidade da posição para leitura na árvore é dificultada pela alta densidade do povoamento ou condições ambientais adversas (LEITE et al., 2011;

BINOTI et al., 2013). Além disso, quando as árvores dominantes não são marcadas na parcela, sendo classificadas como dominantes, árvores diferentes em medições sucessivas, pode haver inconsistência nos dados de crescimento em altura.

Também merece destaque o fato de que a altura é o principal componente do custo de medição de uma parcela permanente em povoamentos de eucalipto (LEITE et al., 2011) e o uso da altura dominante faz com que um número maior de árvores tenham a altura medida na parcela.

Outro inconveniente de realizar a classificação por meio da altura dominante é pressupor que todos os fatores que interferem no crescimento vegetal possam ser representados por uma única variável. Em geral, quando múltiplos fatores são considerados, melhores resultados são obtidos (BRAGA et al., 1999, SCHNEIDER e SCHNEIDER, 2008). Por essa razão, há uma tendência de classificar sítios utilizando um sistema multifatorial (SCHNEIDER e SCHNEIDER, 2008), isto é, correlacionar simultaneamente vários fatores influentes no sítio como solo, precipitação, tendência de crescimento, dentre outros.

Porém, a obtenção de todas essas variáveis é um processo oneroso, além de sua complexidade na aplicação (RIBEIRO, 2015). Para contornar o problema do elevado custo de obtenção de dados, uma alternativa é utilizar outras variáveis de povoamento que já são coletadas ou estimadas nos inventários florestais contínuos. Estudos recentes têm utilizado variáveis dendrométricas na classificação da capacidade produtiva, como: diâmetro dominante (LEITE et al., 2011; CARRIJO, 2016), altura total, diâmetro médio (LEITE, 2014), grau de esbeltez (relação entre altura e o diâmetro a 1,30 m do nível do solo), índice de abrangência (relação entre diâmetro de copa e altura), índice de saliência (relação entre o diâmetro da copa e o diâmetro a 1,30 m do nível do solo) e formal de copa (relação entre diâmetro de copa e altura de copa) (CARRIJO, 2016).

Quanto à complexidade de análise, uma alternativa é utilizar ferramentas de Inteligência Computacional na classificação multifatorial de sítio, por serem capazes de utilizar um grande número de variáveis quantitativas e qualitativas simultaneamente (CONSENZA et al., 2015).

Diante do exposto, uma técnica que pode ser empregada na classificação de povoamentos florestais são as redes neurais artificiais (RNA). Nesse caso, as redes neurais auto-organizáveis são as mais adequadas. Uma representante dessa classe é a rede auto-organizável de Kohonen, sendo a mais utilizada para agrupamento de dados (COSTA et al., 2011).

É importante salientar que na literatura há um número extremamente reduzido de trabalhos que se dedicaram a explorar o uso de RNA para essa finalidade, o que fundamenta a realização de estudos sobre a aplicabilidade dessa ferramenta nesse campo. Nesse sentido, este trabalho analisa o uso de RNA para classificação da capacidade produtiva empregando variáveis dendrométricas.

2 OBJETIVOS

2.1 Objetivo geral

O objetivo geral do trabalho foi avaliar a eficiência da classificação da capacidade produtiva de povoamentos florestais de eucalipto (*Eucalyptus* spp.) por meio de rede neural artificial (RNA).

2.2 Objetivos específicos

Os objetivos específicos foram:

- a) classificar a capacidade produtiva de povoamentos florestais utilizando o método da curva-guia;
- b) classificar a capacidade produtiva de povoamentos florestais utilizando o método da predição dos parâmetros;
- c) classificar a capacidade produtiva de povoamentos florestais utilizando o método da equação das diferenças;
- d) classificar o padrão de crescimento em altura dominante em anamórfico ou polimórfico;
- e) testar o uso das variáveis volume total com casca, área basal, altura total, altura dominante, diâmetro quadrático médio e número de árvores por hectare para classificar a capacidade produtiva de povoamentos florestais por meio de RNA;
- f) selecionar as variáveis dendrométricas que apresentaram os melhores resultados para classificar a capacidade produtiva de povoamentos florestais por meio de RNA;
- g) comparar as classificações obtidas por meio dos diferentes métodos empregados.

3 REVISÃO DE LITERATURA

3.1 Capacidade produtiva de sítios florestais

O sítio é uma fração do ecossistema florestal com características de crescimento semelhantes (BRUM, 1979), determinadas pela combinação de fatores ambientais como edáfico, climático e relevo; e biológicos como a variabilidade genética, a densidade do povoamento, a competição entre plantas e a intensidade de doenças e pragas (GONÇALVES et al., 1990).

A soma de todos os fatores que afetam o crescimento do povoamento florestal representa a capacidade produtiva do sítio (MENEGOL, 1991), definida por Campos e Leite (2013) como o potencial do local para produzir madeira ou outro produto, para determinada espécie ou clone.

De acordo com Skovsgaard e Vanclay (2008) o potencial para produção de biomassa vegetal abrange dois conceitos: o potencial do sítio e a produtividade do sítio. O site potencial é a capacidade do sítio para produzir biomassa vegetal, independente de quanto desse potencial é utilizado pela vegetação. Já o termo produtividade do sítio é frequentemente usado em um sentido mais restrito para se referir à parte do potencial do local que é ou se espera que seja utilizada pelas árvores para a produção de biomassa. Nesse sentido, a produtividade do sítio florestal é definida como a produção que pode ser obtida em um determinado sítio com um determinado genótipo e um regime de manejo específico.

3.2 Por que classificar a capacidade produtiva

O conhecimento da capacidade produtiva é essencial ao bom desenvolvimento de qualquer empreendimento florestal, sendo um fator que influencia no planejamento e na execução das atividades florestais (MENEGOL, 1991).

Um exemplo é a idade de estagnação de um povoamento que varia com o sítio. Geralmente é inversamente proporcional à qualidade do local, sendo atingida em idades menores para sítios de qualidade superior e idades maiores para sítios de qualidade inferior (CALDEIRA et al., 1996). Isso influencia a idade de realização de tratos silviculturais como o desbaste e a colheita (CARVALHO et al., 1999), portanto, é um fator determinante no planejamento das operações de exploração florestal.

Outra atividade florestal que pode ser citada é o controle de plantas daninhas, que geralmente é realizado com maior frequência nos sítios de qualidade superior (SCHNEIDER e SCHNEIDER, 2008).

A classificação da capacidade produtiva também é importante para estimar a produção para longos prazos, pois a atividade florestal apresenta longo ciclo de produção e as empresas necessitam dessas informações previamente para realizar análises econômicas e planejar as operações florestais. Essas estimativas são feitas utilizando modelos de crescimento e produção que comumente requerem uma medida da capacidade produtiva como variável independente (OLIVEIRA, 1998).

A qualidade do sítio é dinâmica e, frequentemente, alguns fatores do sítio podem ser influenciados pelo manejo ao longo do tempo (CALDEIRA et al., 1996) por meio do melhoramento de alguns fatores como nutrientes e densidade do solo, dentre outros. Portanto, o conhecimento da qualidade do sítio auxilia o manejador a determinar as prescrições mais adequadas às condições de cada unidade de manejo. Assim, até mesmo os custos de produção podem variar segundo a qualidade do local (SCHNEIDER e SCHNEIDER, 2008).

3.3 Métodos de classificação da capacidade produtiva

Na literatura são encontrados diversos métodos para classificar a capacidade produtiva do local. Esses métodos avaliam a qualidade do sítio por meio de fatores que apresentem alta correlação com o crescimento vegetal (método indireto) ou por meio de medições diretas do crescimento (método direto) (BRUM, 1979).

Os métodos diretos de classificação podem ser baseados no histórico de produção do local, em dados de volume ou altura dominante, dentre outras variáveis dendrométricas. Já os métodos indiretos podem utilizar relações entre padrões de crescimento de diferentes espécies, características da vegetação natural ou fatores ambientais como relevo, clima e solo (CLUTTER et al., 1983).

A classificação por meio de fatores ambientais e da vegetação indicadora, nem sempre é prática, acessível ou suficientemente precisa para aplicações de manejo (SKOVSGAARD e VANCLAY, 2008). Assim, a classificação por meio de variáveis do povoamento é a mais utilizada e geralmente uma variável ou uma combinação de variáveis da árvore ou do povoamento facilmente medidas são usadas para indicar a produtividade do local

(SKOVSGAARD e VANCLAY, 2008). Nesse tipo de classificação considera-se que o crescimento da vegetação possa refletir todas as interações dos fatores de sítio (TONINI et al., 2001).

Para Clutter et al. (1983) os métodos diretos fornecem melhores avaliações da qualidade do local que os métodos indiretos. Segundo esses autores os métodos indiretos podem ser empregados quando medições da espécie de interesse não estão disponíveis para o local.

Um dos fatores que determinam a escolha do método é a escala de estudo, que pode ser realizada em um aspecto amplo ou específico. A classificação sob o aspecto amplo é realizada para grandes regiões, enquanto a classificação sob o aspecto específico é realizada em uma escala menor, por exemplo, em nível da área de uma empresa, sendo mais apurada que a anterior (SCHNEIDER e SCHNEIDER, 2008). As classificações amplas indicam as potencialidades florestais e podem ser realizadas utilizando o fator clima, por exemplo, já as classificações específicas indicam a capacidade produtiva da propriedade florestal (SCOLFORO, 2006) e podem utilizar, por exemplo, o mapeamento das classes de solo.

Nesse contexto é possível perceber que a escolha do método de classificação também depende do objetivo. Oliveira (1998) exemplifica que a avaliação com base em fatores ambientais pode ser utilizada para a alocação de espécies, a estratificação do terreno, a definição da adubação, definição de unidades de manejo e práticas silviculturais. Já quando o objetivo da avaliação for a prognose da produção podem-se empregar os métodos que utilizam a relação altura dominante e idade.

Também é possível integrar métodos de classificação de forma complementar. Um exemplo disso é realizar o mapeamento da qualidade do local a partir de fatores edáficos para definir unidades de manejo e, posteriormente, avaliar a qualidade de cada unidade por meio da relação altura dominante e idade (OLIVEIRA, 1998).

Uma limitação dos métodos de classificação ocorre devido à diversidade de fatores que atuam conjuntamente na qualidade do sítio, sendo um grande desafio incluir todos esses fatores na avaliação do potencial produtivo do local (TONINI et al., 2001).

Segundo Schneider e Schneider (2008) quanto maior o número de fatores analisados, maior será a precisão e a especificidade da classificação. Esses autores afirmam que a tendência atual é utilizar cada vez mais o sistema multifatorial de classificação do sítio. No entanto, o

número de fatores analisados está condicionado por aspectos como existência de pessoal especializado, finalidade do trabalho, existência de infraestrutura e custos de determinação.

3.3.1 Classificação pelo índice de local

De acordo com Skovsgaard e Vanclay (2008) com a introdução de métodos científicos na silvicultura na Europa há 200-300 anos, as primeiras tentativas de avaliar e classificar o potencial de produção dos sítios florestais foram baseadas nas propriedades do local, incluindo as características do clima, relevo ou solo. Posteriormente, as classes de locais foram classificadas pelo volume do povoamento. No final do século XIX, percebeu-se que a altura média do povoamento, em determinada idade, é uma medida prática da produtividade do local e foi introduzida uma classificação baseada em espécies e padrões de desenvolvimento de altura do sítio.

A classificação do local pela altura do povoamento tornou-se uma das práticas mais universais em silvicultura e é reconhecida como um dos indicadores mais adequados de produtividade do local para fins de manejo em povoamentos florestais equiâneos (SKOVSGAARD e VANCLAY, 2008).

A altura dominante tem sido preferida como indicadora da qualidade do local, pois, dentro de certos limites, não é influenciada pela densidade do povoamento e intervenções silviculturais (GARCIA, 1995). Além disso, essa variável apresenta melhor correlação com as condições ambientais, maior praticidade na aplicação (BRUM, 1979) e para muitas espécies existe uma correlação positiva entre a altura dominante e a produção volumétrica (CLUTTER et al., 1983).

Nesse tipo de classificação, a avaliação da qualidade do sítio é feita por meio do índice de local, isto é, a altura dominante média do povoamento em uma idade específica, denominada idade índice (CAMPOS e LEITE, 2013). A idade índice é arbitrária, mas recomenda-se utilizar uma idade índice próxima à idade de rotação (SCOLFORO, 2006; CAMPOS e LEITE, 2013).

Os índices de local podem ser associados às classes de local. O número de classes de sítio está diretamente relacionado com o intervalo pretendido e a amplitude dos valores observados na idade índice (MARCOLIN, 1990). No entanto, tornou-se habitual adotar três classes de capacidade produtiva (superior, média e inferior).

Embora a produtividade real do local não seja fielmente representada pelo índice de local, esse método é o mais amplamente aceito e provavelmente o mais simples para estimar a produtividade do local (SHARMA et al., 2002).

3.3.1.1 Altura dominante

A altura dominante de um povoamento pode ser determinada a partir de diferentes definições (SCHNEIDER e SCHNEIDER, 2008). Embora essa variável seja amplamente utilizada como medida da qualidade do local não foi adotada uma definição universalmente aceita de altura dominante para estimar o índice de local (SHARMA et al., 2002).

O conceito de altura dominante mais utilizado é o de Assmann (1970), segundo o qual a altura dominante corresponde a altura média de 100 árvores mais grossas por hectare. Uma razão que justifica o uso frequente desse conceito é que no inventário florestal mede-se o diâmetro de todas as árvores da parcela, facilitando assim a identificação das árvores que possuem os maiores diâmetros. Ao utilizar esse conceito, pressupõe-se que as árvores com maior crescimento em diâmetro também apresentam maior incremento em altura.

Para identificação das árvores dominantes devem ser observadas as seguintes características: essas árvores devem ser livres de supressão e danos, a copa deve se situar no dossel superior, devem ser sadias, com copa e fuste bem definidos, sem bifurcação e tortuosidade (OLIVEIRA, 1998).

Girardin e Broquen (1995) estudando o crescimento de duas espécies arbóreas em diferentes condições de sítio determinaram critérios para utilizar a altura dominante como um indicador de condições do local, sendo eles: a população amostrada não deve estar no estado juvenil de crescimento, pois nesse estágio as características de crescimento são muitas vezes variáveis e afetadas por outros fatores que não representam a qualidade do sítio; as densidades não devem ser extremas, devem estar entre 900 e 2500 árvores.ha⁻¹; não tenham sido realizadas práticas silviculturais com remoção de árvores dominantes e as árvores dominantes devem apresentar bom estado fitossanitário e ausência de danos severos.

3.3.1.2 Tipos de curvas de índice de local

As curvas de índice de local são construídas a partir de equações de índices de local e são utilizadas para classificar de forma quantitativa a capacidade produtiva do lugar (CAMPOS e LEITE, 2013).

Por muito tempo o método padrão de construção de curvas de índice de local foram os gráficos anamórficos. Esse método consistia em plotar pares de valores de altura e idade em papel milimetrado e desenhar à mão uma curva-guia, isto é, uma curva média para descrever a tendência dos dados. Todas as outras curvas de índice de local eram então desenhadas proporcionalmente à curva-guia (SUBEDI et al., 2009).

Posteriormente, o método gráfico foi substituído pelo uso de regressão. O procedimento geral é semelhante ao método gráfico, com exceção de que a forma da curva é ditada pela equação selecionada e as curvas são ajustadas pelo método dos mínimos quadrados (SKOVSGAARD e VANCLAY, 2008).

As curvas geradas por esse tipo de método são denominadas anamórficas, pois apresentam inclinação constante, mas variando os interceptos (CLUTTER et al., 1983). Em outras palavras, o parâmetro β_1 do modelo escolhido para descrever a relação altura dominante versus idade é considerado constante para todos os sítios, mas variando o ponto de interseção (β_0). Assim, β_0 é o parâmetro específico do local (SCOLFORO, 2006).

Portanto, gerar curvas de índice de sítio anamórficas é equivalente a assumir que diferenças entre sítios resultam em diferenças similares na taxa de crescimento em altura para todas as idades (MACHADO, 1980). Essas curvas podem ser construídas com dados de altura e idade originados de parcelas permanentes, análise de tronco ou parcelas temporárias (CLUTTER et al., 1983).

O uso de curvas de índice de local anamórficas exige alguns cuidados. Subedi et al. (2009) destacam que se a curva guia for tendenciosa, toda a família de curvas anamórficas será tendenciosa. Os autores também destacam que outro ponto fraco das curvas anamórficas é pressupor uma forma comum para todas as classes de local, já que para muitas espécies a tendência de crescimento em altura varia com a qualidade do local.

A ocorrência de diferenciação na tendência de crescimento em altura em diferentes sítios, para uma mesma espécie, é denominada polimorfismo (SELLE et al., 1994). Para construir curvas polimórficas, são necessárias medições repetidas em intervalos de tempo sucessivos.

Portanto, são utilizados dados de parcelas permanentes ou de análise de tronco de espécies que apresentam anéis de crescimento visíveis (CLUTTER et al., 1983).

Para as curvas polimórficas o parâmetro específico do sítio é a inclinação da curva (β_1), que representa uma taxa de crescimento variável para cada sítio, enquanto o intercepto (β_0) é constante para todos os sítios (SCOLFORO, 2006).

Quanto à influência nas estimativas volumétricas, Scavinski (2014) estudando o efeito de curvas de sítio anamórficas e polimórficas na modelagem do crescimento e da produção em nível de povoamento em plantios de *Eucalyptus grandis*, concluiu que as estimativas de produção utilizando o índice de local obtido tanto na forma anamórfica quanto na forma polimórfica não diferiram estatisticamente dos valores de produção observados.

3.3.1.3 Construção de curvas de índice de local

Há diversos métodos para construção de curvas de índice de local. Alguns dos principais métodos para construção de curva de índice de local são (CAMPOS e LEITE, 2013): método da curva-guia, método da atribuição preliminar de índices de local, método da equação das diferenças e método da predição dos parâmetros.

O método da curva-guia é baseado nos primeiros métodos gráficos (CLUTTER et al., 1983), sendo atualmente o método mais difundido para construção de curvas de índice de local no Brasil (CAMPOS e LEITE, 2013). Os dados podem ser oriundos de parcelas permanentes, parcelas temporárias ou análise de tronco (CLUTTER et al., 1983). As curvas de índice de local obtidas são do tipo anamórficas.

No método da atribuição preliminar de índices de local, o índice é obtido em função das variáveis altura dominante, idade e índices de local definidos preliminarmente, sendo necessários dados de parcelas permanentes ou análise completa do tronco (CAMPOS e LEITE, 2013). Esse método produz curvas anamórficas de índice de local.

Para construção de curvas de índice de local pelo método da equação das diferenças, são necessários dados de parcelas permanentes ou análise de tronco. A equação de índice de local é obtida pela substituição de um parâmetro no modelo de crescimento em altura dominante. A substituição da assíntota produz curvas anamórficas, enquanto a substituição de qualquer outro parâmetro produz curvas polimórficas (ELFVING e KIVISTE, 1997).

O método da predição dos parâmetros produz curvas polimórficas e requer dados de parcelas permanentes ou análise de tronco (CLUTTER et al., 1983).

Dias et al. (2005) comparando a eficiência dos métodos da curva-guia, equação das diferenças e predição de parâmetros para construir curvas de índices de local em povoamentos desbastados de eucalipto, concluíram que todos os métodos foram eficientes segundo os critérios de avaliação adotados. No entanto, os autores destacam que o método da curva-guia foi considerado o mais adequado para construção das curvas de índices de local para a situação estudada.

Oliveira (1998) comparou estimativas de índice de local obtidas pelo método da curva-guia e pelo método da atribuição dos índices de local preliminarmente com valores observados de índice de local (obtidos a partir de medições sucessivas). De acordo com os critérios adotados, as estimativas foram precisas para todos os modelos testados utilizando os diferentes métodos de classificação. No entanto, foi selecionado o método da curva-guia para realizar as estimativas de índice de local.

A construção de curvas de índice de local requer alguns cuidados. Ao utilizar equações para expressar o crescimento em altura das árvores é essencial analisar se a tendência de crescimento para cada local é semelhante ou diferente. A diferenciação indica a necessidade de ajustar equações independentes para cada local, conduzindo dessa forma a uma classificação mais precisa (MARCOLIN, 1990). De acordo com Campos e Leite (2013) um dos principais erros ao classificar a capacidade produtiva por meio de índices de local é usar curvas anamórficas em povoamentos onde a tendência de crescimento é diferenciada por classe de local.

Outra questão importante na construção de curvas de índices de local é a diversidade de genótipos. Oliveira et al. (2008) destacam que com o advento dos plantios clonais de eucalipto, tem-se a necessidade de agrupamento de clones com padrões de desenvolvimento semelhantes, sob um mesmo regime de manejo e espaçamento para construção de curvas de índice de local. Nesse sentido, esses autores propuseram um método para estratificação de florestas de eucalipto visando a construção de curvas de índices de local. O método proposto foi baseado na tendência de crescimento dos genótipos e foi eficiente quando comparado a não estratificação, pois proporcionou estimativas de volume mais precisas próximo à idade de corte.

Tonini et al. (2004) encontraram variação no padrão de crescimento em altura dominante entre clones e inclusive para um mesmo genótipo de *Eucalyptus saligna*. De acordo com os

autores, para o local do estudo, o solo foi o fator ambiental responsável por essa variação. Nessas condições, os autores sugerem que curvas de índice de sítio devem ser construídas observando-se o clone e o grupo de solo.

3.3.1.4 Estabilidade das curvas de índice de local

A classificação pelo índice de local é independente da idade (CAMPOS e LEITE, 2013), isso quer dizer que para uma determinada curva de índice de local, o índice deve ser constante para os diferentes pares de altura dominante e idade.

O procedimento para a avaliação da estabilidade consiste em avaliar a constância da classificação dos sítios. Figueiredo (2005) comparando métodos e modelos para construção de curvas de índice de local para *Tectona grandis*, avaliou a estabilidade de equações de índice de local por meio do desvio padrão das classificações de sítio nas diversas idades para cada parcela. De acordo com o autor, a equação mais estável é aquela que apresenta um maior número de parcelas com desvio padrão das classificações igual a zero, isto é, estabilidade total. Nesse caso, a parcela é classificada em uma mesma classe de sítio do primeiro ao último ano observado. Assim, quanto maior o número de parcelas com estabilidade total, melhor será a equação de classificação de sítio.

No entanto, é importante diferenciar estabilidade das curvas de índice de local e estabilidade da capacidade produtiva. Para uma determinada curva de índice de local, a altura dominante na idade índice é fixa. Assim, para qualquer par de altura e idade, o índice de local obtido pela equação deve ser o mesmo. Portanto, deve haver estabilidade das curvas de índice de local.

Já para a capacidade produtiva, a estabilidade nem sempre se aplica. Skovsgaard e Vancley (2008) destacam que estudos evidenciam que o potencial e a produtividade do sítio florestal não são constantes, mas mudam ao longo do tempo. A capacidade pode ser alterada pelo manejo ou por fatores não controlados como precipitação, por exemplo.

3.3.1.5 Testes de anamorfismo

Após realizar as estimativas de índice de local, pode ser necessário identificar se as curvas são anamórficas ou polimórficas. Scolforo (2006) apresenta dois procedimentos utilizados para essa finalidade.

O primeiro consiste na análise do coeficiente de variação das alturas médias das árvores dominantes por classe de sítio e idade, já que uma das pressuposições básicas do anamorfismo é que este valor seja semelhante em todas as idades nas classes de sítio.

O segundo método verifica a existência de relação linear entre o índice de sítio e as alturas dominantes médias nas várias idades consideradas. Esse método consiste em ajustar um modelo linear simples relacionando as variáveis índice de local e altura dominante média para cada idade. Se o comportamento das curvas for anamórfico, a interseção da reta ajustada será igual a 0 e a inclinação igual a 1 na idade índice. Para as demais classes de idade a inclinação será maior que 1 para idades inferiores à idade índice e menor que 1 para idades superiores à idade índice.

3.4 Novas ferramentas para classificação da capacidade produtiva

Algumas técnicas vêm sendo utilizadas recentemente para classificar a capacidade produtiva de sítios florestais como geoestatística, Light Detection and Ranging (LiDAR), Máquina Vetor de Suporte (MVS) e redes neurais artificiais (RNA).

Pelissari et al. (2014) aplicaram técnicas de modelagem geoestatística para estimar a altura dominante e mapear classes de índice de local em povoamentos de teca no Mato Grosso. Ribeiro (2015) utilizou a krigagem para mapeamento e classificação de sítios de pinus no Paraná.

Martins (2015) utilizou um levantamento LiDAR para estimar a altura de árvores dominantes em um povoamento de eucalipto, visando a construção de curvas de índice de sítio.

Cosenza et al. (2015) utilizaram RNA dos tipos Perceptron de Multicamadas e Função de Base Radial; e MVS para classificação da capacidade produtiva de povoamentos florestais de eucalipto, por meio de variáveis edáficas, de manejo e do povoamento.

3.5 Redes neurais artificiais

As redes neurais artificiais (RNA) são sistemas computacionais constituídos de unidades de processamento distribuídas paralelamente, denominadas neurônios. Os neurônios apresentam uma função de ativação que recebe os sinais de entrada ponderados por conexões, os pesos, que representam o aprendizado da rede (HAYKIN, 2001; BRAGA et al., 2007).

As RNA são inspiradas nas redes neurais biológicas. O cérebro humano é uma rede de aproximadamente 100 bilhões de neurônios (RIBEIRO, 2009) que se comunicam via conexões

sinápticas. Um neurônio biológico é constituído pelos dendritos, o corpo celular e o axônio (Figura 1-A).

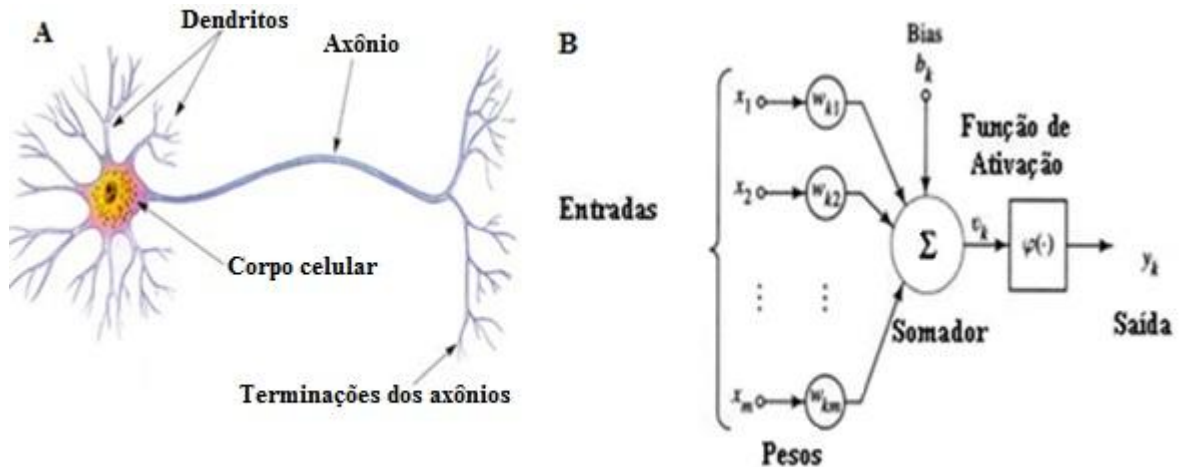


Figura 1- Neurônio biológico (A) e neurônio artificial (B). Fonte: (A) RIBEIRO, 2009; (B) HAYKIN, 2001.

Os impulsos são transmitidos entre neurônios através de sinapses. Uma sinapse é uma conexão que une os neurônios funcionalmente, formando as redes neurais biológicas. O impulso chega ao neurônio receptivo por meio dos dendritos, as chamadas zonas receptivas. No corpo celular os impulsos são combinados e, se o valor resultante for superior ao limiar de excitação do neurônio, é produzido um impulso que é transmitido para as células seguintes por meio do axônio (HAYKIN, 2001; BRAGA et al., 2007).

Analogamente um neurônio artificial (Figura 1-B) apresenta três elementos básicos, um conjunto de sinapses ou elos de conexões, um somador e uma função de ativação. O neurônio artificial funciona da seguinte forma, um sinal de entrada x_j (representando um impulso) na sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico w_{kj} . No somador (representando o corpo celular) é feito o somatório dos sinais de entrada ponderados pelos pesos sinápticos. O valor resultante é submetido a uma função de ativação que gera uma saída (representando o axônio). O *bias* é um parâmetro externo que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação (HAYKIN, 2001).

Os neurônios artificiais individuais apresentam capacidade computacional limitada. No entanto, quando interligados na forma de uma rede, possuem alto poder computacional sendo

capaz de resolver problemas de elevada complexidade. De acordo com Braga et al. (2007), as principais tarefas que as RNA podem executar são: classificação, categorização (agrupamento), aproximação, previsão e otimização.

A estrutura ou arquitetura da rede (isto é: número de neurônios, número de camadas e como esses neurônios estão organizados) depende de vários fatores. Sendo eles: a complexidade do problema, dimensão do espaço de entrada, características dinâmicas ou estáticas e conhecimento prévio sobre o problema (BRAGA et al., 2007). A Figura 2 apresenta um exemplo de arquitetura.

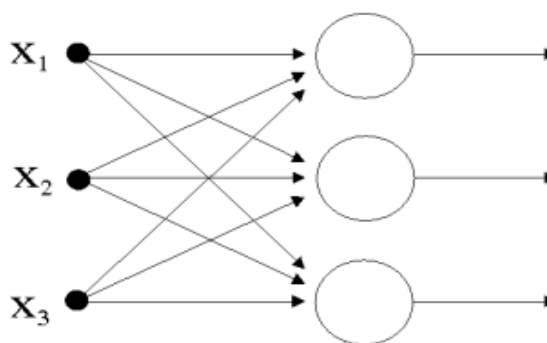


Figura 2- Exemplo de uma arquitetura de RNA de uma única camada. Fonte: http://www.gsigma.ufsc.br/~popov/aulas/rna/uteis/RNA_material_apoio.pdf

A propriedade principal das RNA é a sua capacidade de aprender por meio de exemplos. O conhecimento de uma rede é representado por seus parâmetros, que são os pesos das conexões. O conhecimento é adquirido no treinamento da rede, a cada iteração é feito o ajuste dos pesos de acordo com algum critério pré-estabelecido (HAYKIN, 2001; BRAGA et al., 2007). Há vários algoritmos para treinamento de RNA, podendo ser divididos em dois grupos: treinamento supervisionado e treinamento não supervisionado (BRAGA et al., 2007).

No aprendizado supervisionado o conjunto de treinamento é formado por pares de entrada e saída, ou seja, as variáveis preditoras e a saída desejada. Durante o treinamento a rede calcula uma saída para cada padrão de entrada e ajusta os pesos visando minimizar a distância entre a saída desejada e a saída calculada. O aprendizado supervisionado se aplica a problemas em que se deseja obter um mapeamento entre padrões de entrada e saída.

Já no aprendizado não supervisionado, somente os padrões de entrada são fornecidos à rede. Ao contrário do treinamento supervisionado, não é fornecida a saída desejada explícita ou a

avaliação externa da saída produzida para cada dado de entrada. Dessa forma, o ajuste dos pesos depende apenas dos dados de entrada. Portanto, esse tipo de aprendizado se aplica a problemas que visam extrair características dos dados de entrada como, por exemplo, identificar agrupamentos. Os modelos mais conhecidos de aprendizado não supervisionado são os mapas auto-organizáveis de Kohonen e os modelos *Adaptive Resonance Theory* (ART) (BRAGA et al., 2007).

3.5.1 Mapas auto-organizáveis de Kohonen

Os mapas auto-organizáveis de Kohonen, também conhecidos como redes de Kohonen ou redes SOM (do inglês *Self-Organizing Maps*), foram desenvolvidos pelo professor Teuvo Kohonen na década de 1980 (BRAGA et al., 2007).

Os mapas auto-organizáveis de Kohonen pertencem ao grupo de redes neurais artificiais de treinamento não supervisionado. Isso quer dizer que a única informação fornecida está no conjunto de padrões de entrada e a partir disso a rede define seus parâmetros sem auxílio externo, daí o nome mapas auto-organizáveis (BRAGA et al., 2007).

A auto-organização permite extrair características dos dados (LINDEN, 2009). Essa propriedade é útil em diversas situações em que é necessário o reconhecimento de padrões e agrupamento de dados em que as classes não são previamente conhecidas (BRAGA et al., 2007). De acordo com Silva (2004) a visualização de dados e a análise de agrupamentos são as principais aplicações dos mapas auto-organizáveis.

Essas redes possuem forte inspiração neurofisiológica, pois implementam os conceitos de auto-organização presente no córtex cerebral. O principal resultado da auto-organização cerebral é a formação de um mapa topológico com regiões que apresentam funções específicas (BRAGA et al., 2007). Dessa forma, entradas sensoriais diferentes como a tátil, a visual e a acústica são mapeadas para áreas diferentes do córtex cerebral de uma maneira topologicamente ordenada (HAYKIN, 2001).

A auto-organização da rede é realizada da seguinte forma: quando um vetor de entrada é apresentado, a rede identifica o neurônio que apresenta vetor peso mais semelhante à entrada apresentada. No decorrer do treinamento, a rede diminui a distância entre o vetor peso e o vetor de entrada, incluindo esse ajuste para a vizinhança do neurônio vencedor. Essa última característica permite que a rede construa um mapa topológico em que neurônios que estão

topologicamente próximos respondam de forma semelhante a padrões de entrada semelhantes (BRAGA et al., 2007). Portanto, o mapa auto-organizável identifica, nos padrões de entrada, os agrupamentos e os correlaciona a regiões específicas da grade de neurônios (SILVA, 2004).

A arquitetura básica de uma rede de Kohonen consiste em duas camadas, uma de entrada e outra de saída (competitiva), como mostra a Figura 3. Os neurônios da camada competitiva estão conectados a todos os neurônios da camada de entrada e podem estar dispostos de acordo com uma determinada topologia. A topologia da grade dita como os neurônios estão interconectados por uma relação de vizinhança, podendo ser retangular ou hexagonal. A escolha da topologia e do tamanho do mapa depende da aplicação (COSTA, 1999).

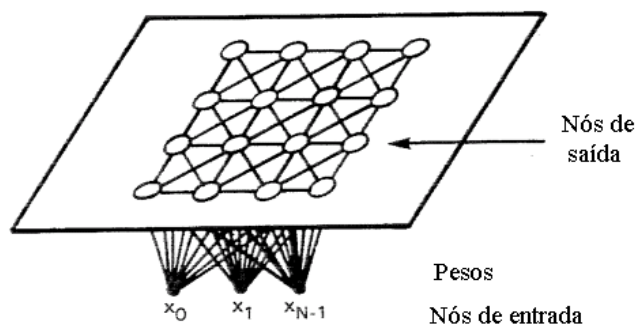


Figura 3- Estrutura básica de uma rede de Kohonen. Fonte: BRAGA et al., 2000.

3.5.1.1 Treinamento sequencial do mapa auto-organizável

A primeira etapa do treinamento é a inicialização. Os pesos iniciais devem ser escolhidos aleatoriamente, o que garante que nenhuma organização prévia seja imposta ao mapa de características (HAYKIN, 2001; BRAGA et al., 2007). Costa (1999) sugere inicializar os vetores com padrões aleatórios do conjunto de dados. Outra forma sugerida pelo autor é utilizar valores máximos e mínimos em cada dimensão. Já Haykin (2001) recomenda que sejam atribuídos valores pequenos tomados de um gerador de números aleatórios.

A próxima etapa é a amostragem, em que padrões de entrada são selecionados aleatoriamente do conjunto de padrões. Posteriormente, três processos básicos se seguem: competição, cooperação e adaptação sináptica (COSTA, 1999; HAYKIN, 2001; BRAGA et al., 2007).

a) Competição

A rede de Kohonen utiliza um algoritmo de aprendizado competitivo (e não supervisionado), em que os neurônios da camada de saída competem entre si para se tornarem ativos.

Suponha que um vetor de entrada seja apresentado à camada de entrada da rede. Cada neurônio da camada de saída irá computar um valor de ativação. O estado de ativação de um neurônio é determinado pela distância entre seu peso e o vetor de entrada. Em geral, a função de ativação é baseada na medida de distância euclidiana (BRAGA et al., 2007):

$$y_j = \sum_{i=1}^n \|x_i - w_{ji}\| \quad (1)$$

em que: y_j é a ativação do neurônio j , w_{ji} é o vetor de peso da conexão entre o elemento de entrada x_i e o neurônio de saída j .

O neurônio vencedor também chamado de BMU (do inglês *Best Match Unit*) é o neurônio que apresenta maior similaridade com o padrão apresentado. Assim, para cada padrão de entrada apenas um neurônio se torna ativo e determina a localização do centro da vizinhança dos neurônios a serem treinados (COSTA, 1999; BRAGA et al., 2007).

b) Cooperação

Sabe-se que no córtex cerebral a ordenação topológica é resultado de interação lateral entre as células. Isso quer dizer que para uma determinada ativação cerebral, o grau de ativação dos neurônios contribui para estimular neurônios próximos e diminui à medida que se aumenta a distância da região da ativação inicial (KOHONEN, 2001).

Em RNA esse fenômeno pode ser modelado por meio de uma função de vizinhança que decresce com o tempo e com a distância em relação ao neurônio vencedor (VESANTO et al., 2003). Por meio da função de vizinhança os vizinhos do neurônio vencedor são selecionados e excitados. Algumas funções que podem ser utilizadas no algoritmo de treinamento para essa finalidade são a função chapéu-mexicano (Figura 4) (COSTA, 1999) e a função Gaussiana, sendo a última a mais utilizada (VESANTO et al., 2003).

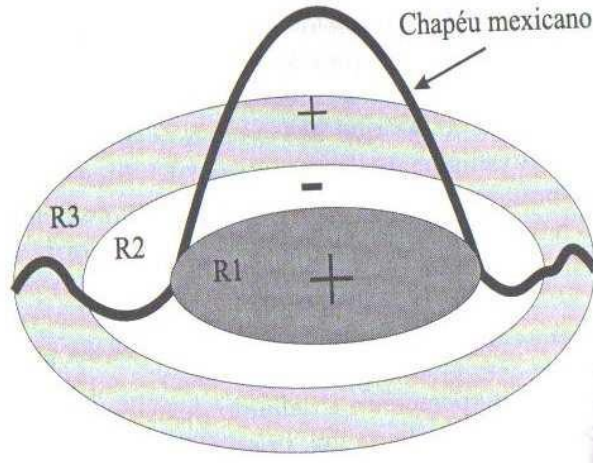


Figura 4 - Tipos de estímulos produzidos de acordo com a função chapéu mexicano, para efeito de ajuste dos pesos sinápticos. Cada neurônio influencia o estado de ativação de seus neurônios vizinhos de três formas possíveis: excitatória (R1), inibitória (R2) e levemente excitatória (R3). Fonte: BRAGA et al., 2007.

A vizinhança do neurônio vencedor pode assumir vários formatos: quadrado, hexágono e círculo. A definição de um formato mais adequado depende do problema atacado e da distribuição dos dados. Geralmente o formato é definido por um processo de tentativa e erro (BRAGA et al., 2007).

c) Adaptação Sináptica

Nesta fase os neurônios excitados ajustam seus pesos sinápticos em relação ao vetor de entrada (HAYKIN, 2001). Não apenas o neurônio vencedor, mas também todos os neurônios localizados em sua vizinhança são ajustados, de acordo com uma função. A equação 2 mostra como são atualizados os pesos da unidade vencedora e de sua vizinhança (BRAGA et al., 2007):

$$w_{ji}(t+1) = \begin{cases} w_{ji}(t) + \eta(t)(x_i(t) - w_{ji}(t)), & \text{se } j \in \Lambda(t) \\ w_{ji}(t), & \text{caso contrário} \end{cases} \quad (2)$$

em que: $w_{ji}(t)$ é o peso da conexão entre o elemento de entrada $x_i(t)$ e o neurônio j , $\eta(t)$ é a taxa de aprendizado e Λ é a vizinhança do neurônio vencedor, tudo no tempo t .

Como mostra a Figura 5, essa regra tem o efeito de mover o vetor de peso sináptico do neurônio vencedor e de sua vizinhança, em direção ao padrão de entrada (HAYKIN, 2001).

Portanto, após várias apresentações do conjunto de treinamento, os vetores de pesos tendem a seguir as distribuições dos vetores de entrada (BRAGA et al., 2007).

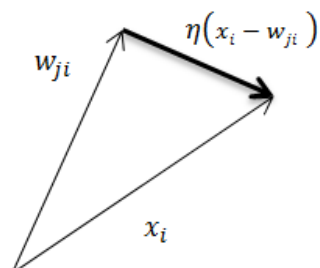


Figura 5- Esquema de atualização dos pesos em uma rede neural competitiva. Em que w_{ji} é o vetor de peso da conexão entre o elemento de entrada x_i e o neurônio de saída j , η é a taxa de aprendizado. Fonte: COSTA, 1999. Adaptado.

Durante o treinamento, a taxa de aprendizado (η) e a vizinhança (λ) são continuamente decrementadas. Esses parâmetros devem assumir um valor pré-fixado e devem decair com o tempo t .

A adaptação sináptica pode ser dividida em duas fases (HAYKIN, 2001; BRAGA et al., 2007): orientação e convergência. A orientação é um mapeamento grosseiro dos padrões de entrada, identificando o número de grupos a ser formado e suas posições no mapa. A taxa de aprendizado é alta, em torno de 1 e reduzida a um valor próximo de 0,1; ocorrendo grandes mudanças nos pesos. A vizinhança pode envolver quase todos os neurônios da rede e ser reduzida até atingir o raio de um ou dois vizinhos. A fase de convergência exige um número maior de ciclos que a etapa anterior, é utilizada uma taxa de aprendizado mais baixa, em torno de 0,01 ou menos, o raio da vizinhança envolve um ou nenhum vizinho. Essa fase sofisticada o mapeamento realizado no estágio anterior, faz um ajuste mais fino do mapa, aprimorando o agrupamento realizado.

Ao final do treinamento, espera-se que entradas semelhantes sejam mapeadas no mesmo neurônio ou em neurônios vizinhos. No entanto, o inverso não é verdadeiro, neurônios vizinhos no mapa podem ser muito distantes no espaço de entrada (COSTA, 1999).

A resposta de interesse do mapa treinado, dependendo da aplicação, pode ser tanto a posição do neurônio vencedor na grade, como o vetor de peso sináptico que está mais próximo do vetor de entrada (HAYKIN, 2001).

3.5.1.2 Visualização do mapa treinado

Após o processo de aprendizagem do mapa é necessário verificar visualmente o resultado da ordenação topológica. Dentre as formas de representação visual destacam-se a matriz-U, os planos de componentes e os histogramas.

3.5.1.2.1 Matriz de distância unificada (matriz-U)

A matriz-U ou matriz de distância unificada é um método de visualização de um mapa treinado, que permite a detecção visual das relações topológicas dos neurônios. A ideia principal é usar a mesma métrica que foi utilizada durante o treinamento para calcular as distâncias entre todos os neurônios vizinhos no arranjo. O resultado obtido a partir da aplicação da matriz-U sobre o mapa é uma imagem em que o nível de intensidade de cada pixel corresponde a uma distância calculada. Valores altos correspondem a neurônios vizinhos dissimilares e valores baixos correspondem a neurônios vizinhos similares (COSTA, 1999).

Esta técnica é extremamente útil quando os vetores de entrada possuem dimensão maior que três. Nesses casos a representação tradicional do mapa, geralmente bidimensional, não é adequada. Pelo fato de geralmente a matriz-U produzir uma imagem relativamente complexa, seu uso é restrito à visualização, sendo uma técnica de auxílio na separação manual dos agrupamentos de um mapa. Nesse caso, o usuário realiza a separação dos grupos baseados nos critérios que lhe pareçam mais adequados (COSTA, 1999).

No entanto, devido à dimensionalidade do conjunto de dados nem sempre é possível obter uma visualização de identificação rápida dos agrupamentos.

A Figura 6 mostra a matriz-U, onde é possível visualizar as classes 1, 2 e 3 referente aos três agrupamentos.

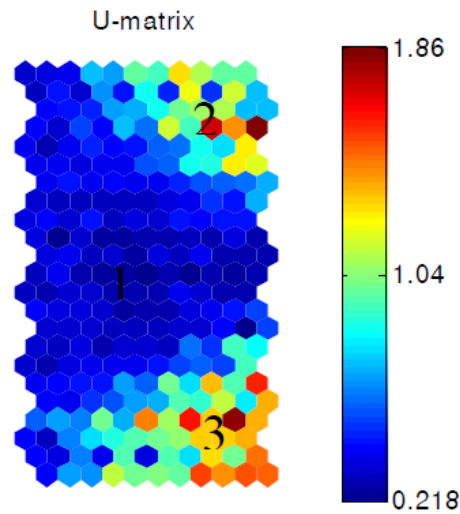


Figura 6- Matriz-U. Valores altos correspondem a neurônios vizinhos dissimilares e valores baixos correspondem a neurônios vizinhos similares. Fonte: FARIA et al., 2010.

3.5.1.2.2 Planos de componentes

Esse tipo de visualização utiliza os valores de cada componente dos vetores de código para colorir o mapa auto-organizável (Figura 7). Para um dado componente j , de um mapa bidimensional $M \times N$, gera-se uma imagem $f(x, y)$ com dimensões iguais às do mapa $M \times N$, onde cada pixel corresponderá ao valor do componente j na posição (x, y) (SILVA, 2004).

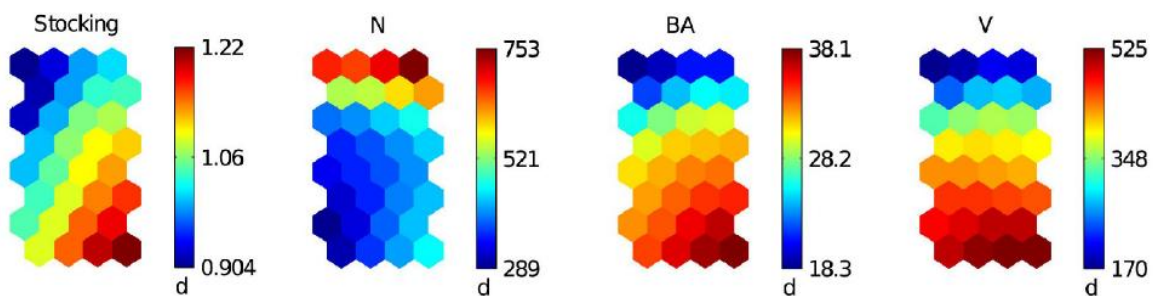


Figura 7- Plano de componentes. Distribuição de cada componente dos vetores de peso sináptico no mapa auto-organizável. Fonte: KLOBUCAR e SUBASIC, 2012.

A análise do plano de componentes permite avaliar a distribuição de cada variável no mapa, portanto, a contribuição de cada variável individualmente para o agrupamento.

Sabe-se que um dos fatores que mais influencia o resultado de uma análise de agrupamentos são as variáveis utilizadas para descrever os objetos. Dessa forma, esse tipo de

visualização é importante na identificação de variáveis que não são discriminatórias e que poderiam ser descartadas da análise. Isso porque variáveis que apresentam praticamente o mesmo valor para todos os objetos terão distribuição uniforme no mapa.

Além disso, após a segmentação do mapa é possível fazer inferências da distribuição das variáveis dentro de cada grupo apenas com a análise visual.

3.5.1.2.3 Histograma

Esse tipo de visualização mostra o nível de atividade dos neurônios da grade de saída (Figura 8). O nível de atividade de um neurônio representa a quantidade de dados do espaço de entrada que são mapeados para este neurônio. Em outras palavras, a frequência em que esse neurônio foi o vencedor no treinamento.

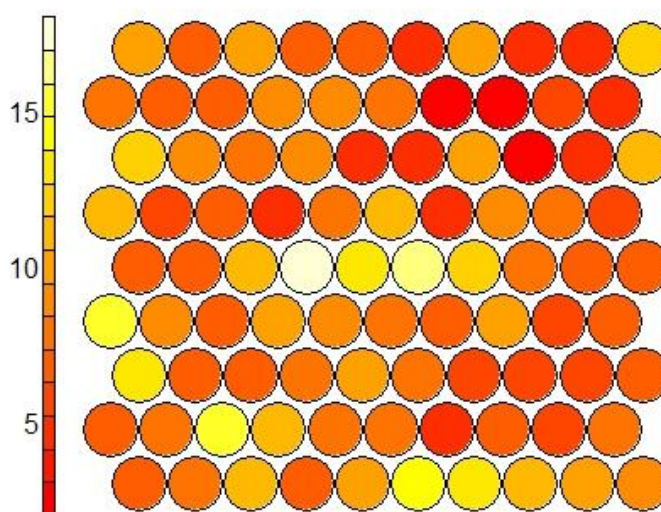


Figura 8- Histograma da frequência de ativação dos neurônios da grade.

3.5.1.3 Medidas de erro

As medidas de qualidade mais comumente usadas são erro de quantização e erro topográfico.

O erro de quantização (E_q) representa a média das distâncias entre cada vetor de dados (x_i) e o correspondente vetor de peso (w_{BMU}) do neurônio vencedor (BMU) no processo competitivo para o padrão (x_i). Este erro varia de 0 a 1, caso a normalização dos vetores de entrada tenha ocorrido nesta mesma grandeza. Estima-se que quanto menor o erro de

quantização, mais bem ajustado o neurônio vencedor estará aos vetores de entrada. O erro é calculado pela equação (3) (FARIA et al., 2010).

$$E_q = \frac{1}{N} \sum_{i=1}^N |x_i - w_{BMU}| \quad (3)$$

O erro topográfico (E_t) avalia o quanto a estrutura da grade aproxima padrões próximos no espaço de entrada. É calculado verificando-se para todas as entradas qual é o neurônio mais bem ajustado (BMU) e também o segundo neurônio mais bem ajustado (BMU2). Este erro será maior quanto maior for a distância entre o primeiro e o segundo neurônio vencedor para a média (FARIA et al., 2010). É dado pela equação:

$$E_t = \frac{1}{N} \sum_{i=1}^N u(x_i) \quad (4)$$

em que $u(x_i)$ corresponde a 1, se o BMU e BMU2 não são vizinhos e 0 caso contrário.

De acordo com Klobucar e Subasic (2012) embora o erro de quantificação forneça informações fáceis de serem interpretadas, sua desvantagem é que ele diminui com o aumento do tamanho do mapa. Isto o torna um pouco inadequado, especialmente quando mapas grandes são considerados. Por outro lado, a medida de erro topográfico não é confiável para mapas pequenos. O seu valor é muito baixo para mapas que não contêm um número alto de neurônios.

3.5.1.4 Segmentação do mapa treinado

De acordo com Costa (1999) o mapa treinado não é suficiente para efetuar análise de agrupamentos, isso porque a informação de distâncias entre os neurônios é perdida. Geralmente a saída de um mapa treinado, para um dado padrão de entrada, é a posição do neurônio vencedor na grade e seu respectivo vetor de peso sináptico (HAYKIN, 2001).

A exceção é a análise simples de agrupamento, em que o número de neurônios é definido como o número de agrupamentos possíveis (ULTSCH, 1993). Curry et al. (2001), por exemplo, utilizando um mapa 2x3 para identificação de grupos estratégicos de empresas de uma rede hoteleira trabalharam com um neurônio por agrupamento. Nesse caso, após o treinamento da

rede, cada neurônio estará associado a um grupo de vetores de entrada. No entanto, este método é aplicado para o caso de redes pequenas (SILVA, 2004).

Santos (2014) estudando métodos de estratificação de florestas de eucalipto visando a amostragem para inventário florestal agrupou os dados utilizando redes auto-organizáveis de Kohonen em que o número de neurônios era igual ao número de grupos.

Em geral, o uso do mapa auto-organizável em análise de agrupamentos requer ferramentas adicionais. Nesse sentido, métodos de pós-processamento do mapa treinado têm sido propostos (COSTA et al., 2011) visando particionar o mapa em grupos de neurônios. A partição da grade em conjuntos de neurônios também determina uma partição no conjunto de dados de entradas, bastando para isso saber para cada dado de entrada seu neurônio de referência na grade (NOLETO, 2007).

Muitos algoritmos para segmentação automática foram desenvolvidos, como o SL-SOM (*Self Labeled SOM*), que realiza a segmentação da matriz-U (COSTA, 1999), a segmentação utilizando técnicas de particionamento de grafos por eliminação de arestas inconsistentes (COSTA e NETTO, 2003), o algoritmo GSOM (GALETE, 2012), dentre outros.

A segmentação do mapa também pode ser realizada por meio de um método tradicional de agrupamento. Vesanto e Alhoniemi (2000b) utilizaram os métodos de agrupamento hierárquico e partitivo para essa finalidade. Esses autores realizaram o agrupamento em dois níveis. Na primeira etapa os dados foram utilizados para treinar a rede e na segunda etapa os vetores de pesos sinápticos foram agrupados utilizando um método hierárquico aglomerativo e o algoritmo k-means. De acordo com os autores, o benefício mais importante deste procedimento (em relação à análise de agrupamento convencional) é que a carga computacional diminui consideravelmente, tornando possível agrupar grandes conjuntos de dados.

Outros exemplos da combinação de mapas auto-organizáveis e um método tradicional de agrupamento são encontrados na literatura. Gonçalves et al. (2008), trabalhando com classificação não-supervisionada de padrões de cobertura terrestre em imagens de satélite utilizaram um método de agrupamento hierárquico com conectividade restrita para segmentação do mapa treinado. Arcoverde et al. (2011) utilizando mapas auto-organizáveis para identificação de áreas prioritárias para a recuperação florestal, realizaram o agrupamento do mapa por meio do algoritmo k-means. Klobucar e Subasic (2012) aplicando o mapa de Kohonen para visualização e

análise de dados de inventário florestal, agruparam os neurônios da rede utilizando o k-means e o método de Ward.

Ao utilizar um método tradicional de agrupamento para segmentação do mapa, o número de grupos deve ser previamente definido pelo usuário. Nessa situação, uma alternativa para definir o número de grupos é analisar o comportamento da variância para uma série crescente de partições. Dessa forma, o número de grupos pode ser estabelecido a partir do patamar de estabilização da variabilidade (RABELO, 2014).

4 MATERIAIS E MÉTODOS

4.1 Descrição dos dados

Os dados utilizados no estudo foram provenientes de inventários florestais contínuos conduzidos em povoamentos de clones de eucalipto (*Eucalyptus* spp.) localizados no estado de Minas Gerais. Esses dados foram cedidos por uma empresa, sendo o processamento em nível de parcela realizado pela própria empresa.

A consistência do banco de dados foi realizada analisando: valores máximos e mínimos para cada variável; o realismo biológico do incremento para as variáveis volume total com casca (V), área basal (B), altura total (Ht), altura dominante (Hd) e diâmetro quadrático médio (q); gráficos de dispersão para a relação entre as variáveis V , B , Ht , Hd , q e idade; a constância da área, genótipo e projeto de cada talhão em suas diferentes parcelas e nas diferentes medições, considerando um único ciclo; a constância da idade para parcelas de um mesmo talhão em cada medição; redundância de informação de parcela ou de talhão.

Após a consistência, foram qualificadas para o estudo 2.334 parcelas distribuídas em 1.703 talhões, compreendendo 60 projetos e 50 genótipos, totalizando uma área de 40.649,11 ha. A densidade do povoamento variou de 259 a 1.602 árvores.ha⁻¹, com um valor médio de 1.067 árvores.ha⁻¹, que corresponde a um espaçamento médio de 9 m².árvore⁻¹.

As medições das parcelas ocorreram aproximadamente nas idades de 36, 48, 60, 72, 84 e 96 meses, entre o período de 2003 a 2011. Algumas parcelas não foram mensuradas em todas as medições. É importante destacar que as parcelas apresentaram diferentes idades dentro de uma mesma medição. A descrição dos dados utilizados é apresentada na Tabela 1.

Tabela 1 - Descrição estatística dos dados utilizados. Em que CV é o coeficiente de variação, V é o volume total com casca, B é a área basal, Ht é altura total, Hd é a altura dominante e q é o diâmetro quadrático médio

Idade (meses)		V (m ³ .ha ⁻¹)	B (m ² .ha ⁻¹)	Ht (m)	Hd (m)	q (cm)
36	Máximo	189,82	20,14	24,6	29,4	15,3
	Médio	73,83	10,04	15,8	17,5	10,8
	Mínimo	10,38	2,98	8,4	9,3	6,3
	CV	36,33	24,14	15,7	14,0	11,0

Continua

Tabela 1 – Continuação

Idade (meses)		<i>V</i> (m ³ .ha ⁻¹)	<i>B</i> (m ² .ha ⁻¹)	<i>Ht</i> (m)	<i>Hd</i> (m)	<i>q</i> (cm)
48	Máximo	265,06	28,26	25,7	30,6	16,8
	Médio	121,50	13,25	19,7	21,8	12,5
	Mínimo	30,57	4,87	12,1	13,7	8,5
	CV	29,81	19,58	12,7	11,2	9,0
60	Máximo	351,29	29,77	29,4	32,0	18,3
	Médio	159,77	15,51	22,1	24,6	13,6
	Mínimo	43,99	6,45	13,6	15,7	9,6
	CV	29,76	18,30	12,8	10,7	8,6
72	Máximo	455,91	33,99	31,3	35,3	21,4
	Médio	187,40	16,84	23,6	26,6	14,3
	Mínimo	62,12	8,32	15,7	18,5	10,1
	CV	30,41	17,23	12,9	10,8	8,5
84	Máximo	408,86	34,33	33,0	36,7	19,5
	Médio	216,60	18,29	25,0	28,5	14,9
	Mínimo	78,78	9,91	16,7	19,9	11,0
	CV	27,83	16,65	13,5	11,0	8,0
96	Máximo	406,44	27,74	32,8	36,7	19,5
	Médio	243,70	19,39	26,3	30,1	15,3
	Mínimo	90,13	11,66	17,4	20,9	11,5
	CV	24,31	13,60	12,0	10,0	7,4

A unidade de manejo considerada para o estudo foi o talhão. Para os talhões que apresentaram mais de uma parcela foi calculada a média para as variáveis altura dominante, altura total, diâmetro quadrático médio, área basal e volume total com casca.

4.2 Classificação da capacidade produtiva

A estratificação do reflorestamento em classes de capacidade produtiva foi realizada para a última medição (2011), pois nela está contida a informação mais atual da floresta. Nessa medição foram amostrados 750 talhões, com idades variando de 36 a 96 meses (Tabela 2).

As classes de capacidade produtiva foram obtidas empregando os seguintes métodos: agrupamento de índices de local obtidos pelo método da curva-guia, que produz curvas anamórficas; agrupamento de índices de local obtidos pelo método da predição de parâmetros, que produz curvas polimórficas; agrupamento de índices de local obtidos pelo método da equação

das diferenças, que produz curvas anamórficas ou polimórficas; e redes neurais auto-organizáveis de Kohonen.

Tabela 2 - Descrição estatística dos dados da última medição (2011). Em que CV é o coeficiente de variação, V é o volume total com casca, B é a área basal, Ht é altura total, Hd é a altura dominante, q é o diâmetro quadrático médio e N é o número de talhões

Idade (meses)	N		V (m ³ .ha ⁻¹)	B (m ² .ha ⁻¹)	Ht (m)	Hd (m)	q (cm)
36	5	Máximo	94,40	11,13	18,0	19,9	11,7
		Médio	74,49	9,91	16,3	17,9	10,9
		Mínimo	49,20	8,16	13,9	15,1	10,1
		CV	26,12	13,21	11,1	11,7	6,4
48	188	Máximo	265,06	28,26	23,2	28,0	14,1
		Médio	126,63	12,97	20,2	22,0	12,1
		Mínimo	54,55	7,65	16,3	17,5	9,5
		CV	21,30	14,90	6,5	7,6	5,8
60	156	Máximo	322,40	21,71	26,7	30,8	15,4
		Médio	181,21	15,70	22,9	25,2	13,2
		Mínimo	92,77	9,88	17,9	19,7	10,9
		CV	20,69	11,95	6,7	7,6	5,6
72	98	Máximo	455,91	26,55	31,3	33,6	21,4
		Médio	257,17	18,88	26,2	28,5	14,7
		Mínimo	132,67	13,17	22,1	24,5	11,9
		CV	28,06	12,00	8,1	7,7	8,9
84	148	Máximo	408,86	28,31	33,0	35,4	17,6
		Médio	230,69	19,03	25,4	29,4	15,2
		Mínimo	117,60	12,38	19,1	24,5	12,5
		CV	25,85	16,32	12,6	8,4	7,3
96	155	Máximo	404,36	27,74	31,1	36,7	18,7
		Médio	247,07	19,61	26,7	30,6	15,4
		Mínimo	127,20	12,90	21,4	25,6	12,9
		CV	19,40	11,74	8,7	7,5	6,1

4.2.1 Métodos da curva-guia, predição dos parâmetros e equação das diferenças

Na classificação utilizando o método da curva-guia foi ajustada a função de crescimento de Schumacher (1939), na forma linear:

$$\ln Hd = \beta_0 + \beta_1 I^{-1} + \varepsilon \quad (5)$$

em que: \ln é o logaritmo com base e , sendo $e \cong 2,718$; Hd é a altura dominante, em metros; I é a idade, em meses; β_0 e β_1 são os parâmetros do modelo; e ε é o erro aleatório, sendo $\varepsilon \sim \text{NID}(0, \sigma^2)$. O ajuste do modelo de Schumacher foi realizado conforme apresentado na Figura 9.

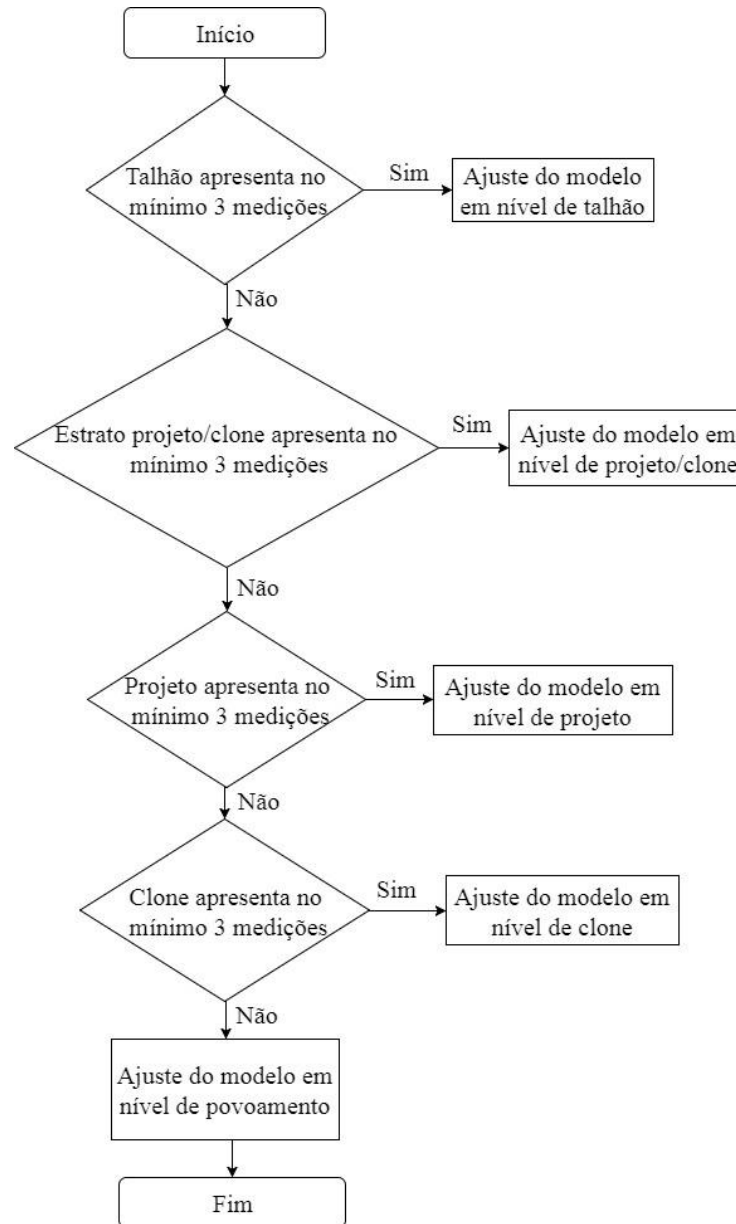


Figura 9 - Fluxograma para ajuste do modelo de Schumacher (1939) linearizado para estimativas de índice de local pelo método da curva-guia.

O índice de local foi obtido pela equação (6) e as curvas de índice de local foram construídas por meio da equação (7) conforme Clutter et al. (1983):

$$Ln(S) = Ln(Hd) - \hat{\beta}_1[I^{-1} - I_i^{-1}] \quad (6)$$

$$Ln(Hd) = Ln(S) + \hat{\beta}_1[I^{-1} - I_i^{-1}] \quad (7)$$

em que: S é o índice de local, em metros; I_i é a idade índice, em meses; $\hat{\beta}_1$ é o parâmetro estimado para o modelo 5; Ln , Hd e I idem modelo 5.

Para o método da predição de parâmetros foi ajustado o modelo logístico para as parcelas que apresentaram no mínimo quatro medições, o número mínimo de observações foi determinado visando representatividade da tendência dos dados:

$$Hd = \beta_0(1 + \beta_1 e^{-\beta_2 I})^{-1} + \varepsilon \quad (8)$$

Os valores iniciais dos parâmetros foram determinados utilizando o método proposto por Camargos (2017):

$$\beta_0 = \text{máximo}(Hd) \quad (9)$$

$$\beta_1 = \frac{\beta_0}{\text{mínimo}(Hd)} - 1 \quad (10)$$

$$\beta_2 = \frac{Ln(\beta_1)}{i} \quad (11)$$

em que i é a idade de inflexão, isto é, a idade em que a taxa de crescimento é máxima, em meses.

Após os ajustes, a idade-índice foi substituída em cada equação ajustada para encontrar o índice de local por parcela:

$$S = \hat{\beta}_0(1 + \hat{\beta}_1 e^{-\hat{\beta}_2 I_i})^{-1} \quad (12)$$

Em seguida, utilizando as estimativas dos parâmetros do modelo (8) e os índices de local obtidos por parcela, foram ajustados os seguintes modelos (CLUTTER et al., 1983):

$$\beta_0 = \alpha_0 + \alpha_1 S + \varepsilon \quad (13)$$

$$\beta_1 = \alpha_2 + \alpha_3 S + \alpha_4 S^2 + \varepsilon \quad (14)$$

$$\beta_2 = \alpha_5 + \alpha_6 S + \alpha_7 S^2 + \varepsilon \quad (15)$$

Esses ajustes substituíram os parâmetros do modelo (7) para estimar o índice de local por talhão e para construção de curvas de índice de local:

$$Hd = (\hat{\alpha}_0 + \hat{\alpha}_1 S) \left(1 + (\hat{\alpha}_2 + \hat{\alpha}_3 S + \hat{\alpha}_4 S^2) e^{-(\hat{\alpha}_5 + \hat{\alpha}_6 S + \hat{\alpha}_7 S^2)I} \right)^{-1} \quad (16)$$

No método da equação das diferenças foi utilizado o modelo de Schumacher (1939) linearizado (modelo 5). Para obter o parâmetro β_1 ajustou-se a seguinte relação funcional (CLUTTER et al., 1983) :

$$Y = \beta_1 X + \varepsilon \quad (17)$$

em que: $Y = \ln(Hd_2) - \ln(Hd_1)$ e $X = I_2^{-1} - I_1^{-1}$.

O ajuste do modelo 17 foi realizado conforme a Figura 9, mas adotando um número mínimo de quatro medições. Obtido o parâmetro $\hat{\beta}_1$ foram realizadas estimativas de índice de local por meio da equação (18) e construídas curvas de índice de local por meio da equação (19):

$$\ln(S) = \ln(Hd) + \hat{\beta}_1 (I_i^{-1} - I^{-1}) \quad (18)$$

$$\ln(Hd) = \ln(S) - \hat{\beta}_1 (I_i^{-1} - I^{-1}) \quad (19)$$

Nos três métodos foi adotada uma idade índice de 72 meses. Obtidos os índices de local para cada talhão por meio dos diferentes métodos, esses índices foram agrupados em três classes de produtividade: superior, média e inferior. A amplitude de classe foi determinada pela diferença entre o valor máximo e o valor mínimo de índice de local dividido por três.

4.2.1.1 Testes de anamorfismo

Foram aplicados dois testes para verificar o anamorfismo ou polimorfismo para as estimativas de índice de local obtidas pelos métodos da curva-guia, predição dos parâmetros e equação das diferenças. Os testes aplicados foram (SCOLFORO, 2006):

a) verificação do coeficiente de variação das alturas das árvores dominantes por classe de índice de local e idade;

b) verificação da existência de relação linear entre os índices de local e as alturas dominantes médias. Para isso foi ajustado o modelo (20) para cada classe de idade:

$$S = \beta_0 + \beta_1 Hd + \varepsilon. \quad (20)$$

4.2.2 Rede neural artificial

Na classificação da capacidade produtiva por meio de rede neural artificial foram utilizadas as variáveis V , B , Ht , Hd e q ; pela relação com o crescimento do povoamento e a qualidade do local (OLIVEIRA et al., 2008); e número de árvores por hectare (Arv_ha) devido à influência dessa variável no crescimento e na produção do povoamento florestal.

O tipo de rede utilizada foi a rede auto-organizável de Kohonen. Para utilizar essa técnica para classificação é necessário que os talhões apresentem a mesma idade. Se essa condição não for atendida, a classe de capacidade produtiva do talhão dependerá da idade e não do potencial de produção. Nesse sentido, foi utilizada a idade de referência empregada nas estimativas de índice de local, ou seja, 72 meses.

Dessa forma, foi necessário estimar as variáveis V , B , Ht , Hd e q para os talhões que não apresentaram medição aos 72 meses. O modelo logístico foi ajustado para cada variável de povoamento:

$$Y = \beta_0 (1 + \beta_1 e^{-\beta_2 I})^{-1} + \varepsilon \quad (21)$$

em que: Y é a variável de povoamento; β_0 , β_1 e β_2 são os parâmetros do modelo; I é a idade em meses e ε é o erro aleatório.

Os valores iniciais dos parâmetros foram determinados utilizando o método proposto por Camargos (2017). O ajuste do modelo logístico foi realizado em nível de talhão, projeto ou

povoamento (Figura 10). Os valores estimados foram corrigidos por meio do índice de consistência (equação 22) tendo em vista que pelo modelo logístico a variável de povoamento é estimada apenas em função da idade e que não foi possível realizar todos os ajustes em nível de talhão. Portanto, o objetivo foi evitar que nos casos em que fosse necessário utilizar uma equação geral, com ajuste em nível de projeto ou povoamento, os talhões apresentassem os mesmos valores estimados para a idade de 72 meses.

$$Y_{72} = \hat{Y}_{72} Y_1 \hat{Y}_1^{-1} \quad (22)$$

em que: Y_{72} é o valor estimado corrigido para a idade de 72 meses; \hat{Y}_{72} é o valor estimado para a idade de 72 meses; Y_1 é o valor observado para a variável de povoamento na idade da última medição e \hat{Y}_1 é o valor estimado para a variável de povoamento na idade da última medição.

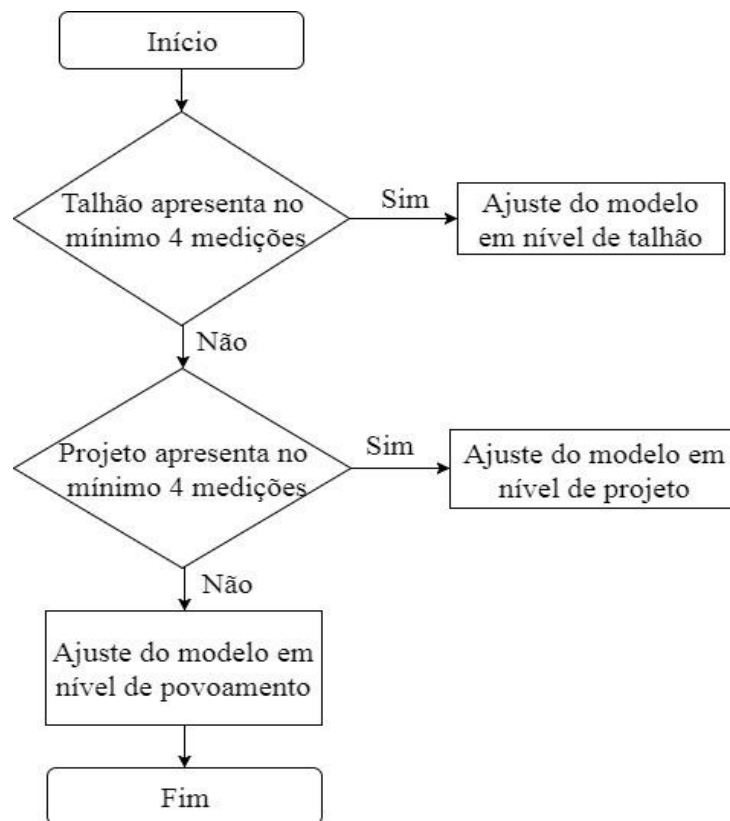


Figura 10 – Fluxograma para ajuste do modelo logístico para estimar as variáveis volume total com casca, área basal, altura total, altura dominante e diâmetro quadrático médio aos 72 meses.

Após obter as estimativas para os 72 meses, foi realizada a padronização das variáveis de entrada da rede, pois apresentavam escalas muito diferentes. O dimensionamento é feito visando obter média zero e variância unitária. Esse procedimento é importante uma vez que o algoritmo de treino da rede usa a distância euclidiana para medir distâncias entre vetores. Assim, a padronização evita que as variáveis de maior escala tenham maior impacto sobre a distância medida (VESANTO et al., 2000a). Com os dados padronizados foram testadas diferentes entradas para as redes neurais, conforme apresentado na Tabela 3.

Tabela 3 – Testes executados para definir a entrada das redes auto-organizáveis de Kohonen. Em que Hd = altura dominante, Ht = altura total, q = diâmetro quadrático médio, B = área basal , V = volume total com casca e Arv_ha = número de árvores por hectare

Teste	Variáveis utilizadas
1	V
2	B
3	Ht
4	Hd
5	q
6	Hd, Ht, q, B, V, Arv_ha

No teste 6 (Tabela 3), primeiramente foi realizada a classificação utilizando todas as variáveis. Posteriormente, foi utilizada a Análise de Componentes Principais (ACP) com o objetivo de selecionar as variáveis por meio de sua contribuição na variação total dos dados. A ACP foi realizada com base na matriz de correlação. Para determinar os componentes principais retidos na análise utilizou-se o critério adotado por Oliveira et al. (2008). Assim, obtida a classificação utilizando todas as variáveis foi determinado o parâmetro de menor importância e o agrupamento foi realizado novamente para comparação. Se o novo agrupamento fosse igual ao agrupamento com todas variáveis, esse parâmetro seria descartado e o processo seria repetido até que houvesse alteração do agrupamento, indicando que a variável deveria ser mantida.

Os parâmetros de treino foram determinados com base na literatura e em valores padrões do pacote utilizado. O número de neurônios das redes foi determinado utilizando uma fórmula heurística proposta por Vesanto et al. (2000a), em que o número de neurônios é obtido por $5\sqrt{n}$, sendo n o número de amostras. No entanto, esse valor poderia ser alterado ao avaliar o histograma de frequência de ativação. Quando a rede apresentava muitos neurônios com

frequência de ativação zero, o tamanho da rede era reduzido (LYNN, 2014). Os demais parâmetros utilizados foram: topologia hexagonal; 100 iterações; taxa de aprendizado inicial e final de 0,05 e 0,01, respectivamente; inicialização aleatoriamente a partir do banco de dados e a medida de dissimilaridade utilizada foi a distância euclidiana.

A resposta da rede foram os vetores de pesos sinápticos que foram agrupados em três grupos. Foi utilizado o método do vizinho mais distante e a medida de dissimilaridade utilizada foi a distância euclidiana. Cada grupo representou uma classe de produtividade: superior, média e inferior.

Foi realizada a análise discriminante para cada teste, visando determinar a entrada que proporcionou maior porcentagem de classificações corretas. A entrada selecionada pela análise discriminante foi utilizada para comparação com os métodos tradicionais.

4.3 Comparação dos métodos

A comparação da classificação pelos diferentes métodos foi realizada por meio da porcentagem de coincidência na alocação dos talhões, área e estoque volumétrico por classe de capacidade produtiva. Uma segunda comparação foi realizada por meio das diferenças percentuais do volume total por classe de sítio em relação ao método da curva-guia. A curva-guia foi considerada padrão por ser o método mais utilizado no Brasil para classificação da capacidade produtiva de povoamentos florestais equiâneos.

4.4 Análises Estatísticas

Os modelos lineares foram ajustados pelo método dos mínimos quadrados ordinários. A avaliação de qualidade dos ajustes foi realizada por meio do coeficiente de determinação, coeficiente de determinação ajustado e do erro-padrão da regressão em porcentagem para os modelos com intercepto e por meio do erro-padrão da regressão em porcentagem e da raiz quadrada do erro médio para os modelos sem intercepto.

Os modelos não lineares foram ajustados por meio do algoritmo NL2SOL (DENNIS et al., 1977) e a avaliação de qualidade dos ajustes foi realizada por meio da raiz quadrada do erro médio e do erro-padrão da regressão em porcentagem.

Na etapa de treinamento das redes neurais artificiais foi utilizado o pacote kohonen (WEHRENS e BUYDENS, 2007) versão 3.0.2.

O processamento foi realizado no *software* R versão 3.3.3 (R CORE TEAM, 2017), com apoio da plataforma RStudio versão 1.0.136 (RSTUDIO TEAM, 2016).

5 RESULTADOS E DISCUSSÃO

5.1 Análise exploratória dos dados

A Figura 11 apresenta o comportamento das variáveis do povoamento florestal. Essas informações são importantes por possibilitar melhor compreensão dos dados utilizados e da relação entre as variáveis.

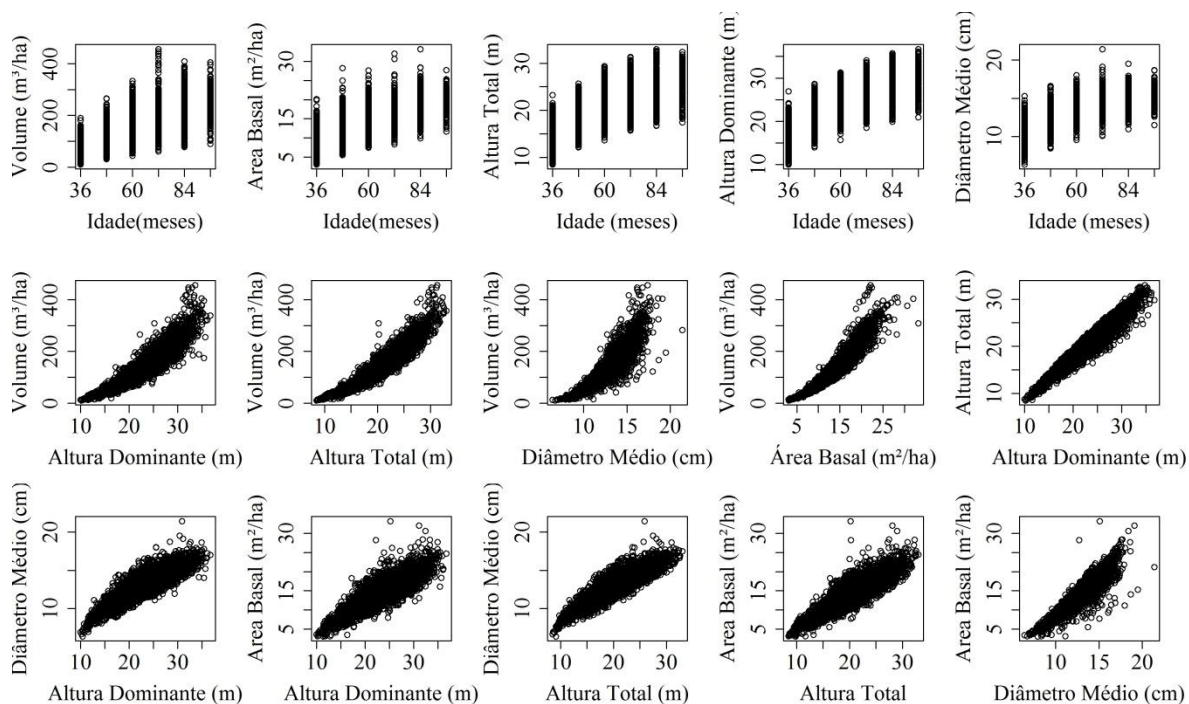


Figura 11 - Dispersão para a relação entre as variáveis volume total com casca, área basal, altura total, altura dominante, diâmetro quadrático médio e idade.

5.1 Método da curva-guia

Os valores médios dos parâmetros estimados para o modelo de Schumacher linearizado foram $\hat{\beta}_0 = 3,715$ e $\hat{\beta}_1 = -29,236$. O resumo dos ajustes é apresentado na Tabela 4.

Analisando a Tabela 4 é possível notar que a precisão aumenta com a redução do estrato. Esse comportamento era esperado, pois há menor variabilidade nos menores estratos. No entanto, devido à insuficiência de dados para ajuste, há situações em que é necessário utilizar equações gerais, sendo essas menos precisas. Os dados de ajuste possibilitaram obter uma equação específica para o talhão em 74% dos casos e em apenas 0,4% do número de talhões foi necessário utilizar uma equação em nível de povoamento.

Tabela 4 – Valores médios para os parâmetros estimados ($\hat{\beta}_0$ e $\hat{\beta}_1$), coeficiente de determinação em porcentagem (R^2) e erro-padrão da regressão em porcentagem ($S_{yx\%}$) para ajustes do modelo de Schumacher (1939) linearizado, em que N é o número de ajustes

Nível	N	$\hat{\beta}_0$	$\hat{\beta}_1$	R^2	$S_{yx\%}$
Talhão	557	3,725	-29,769	97,40	3,18
Projeto/Clone	114	3,676	-27,213	84,29	5,51
Projeto	49	3,673	-27,090	79,62	6,29
Clone	27	3,756	-30,483	81,67	7,57
Povoamento	3	3,712	-31,027	67,60	11,49

As curvas de índice de local anamórficas foram construídas a partir da seguinte equação:

$$\ln(Hd) = \ln(S) - 29,236[(I^{-1}) - (72^{-1})], \quad (23)$$

com variação do índice de local de 22 a 34 m na idade índice de 72 meses (Figura 12).

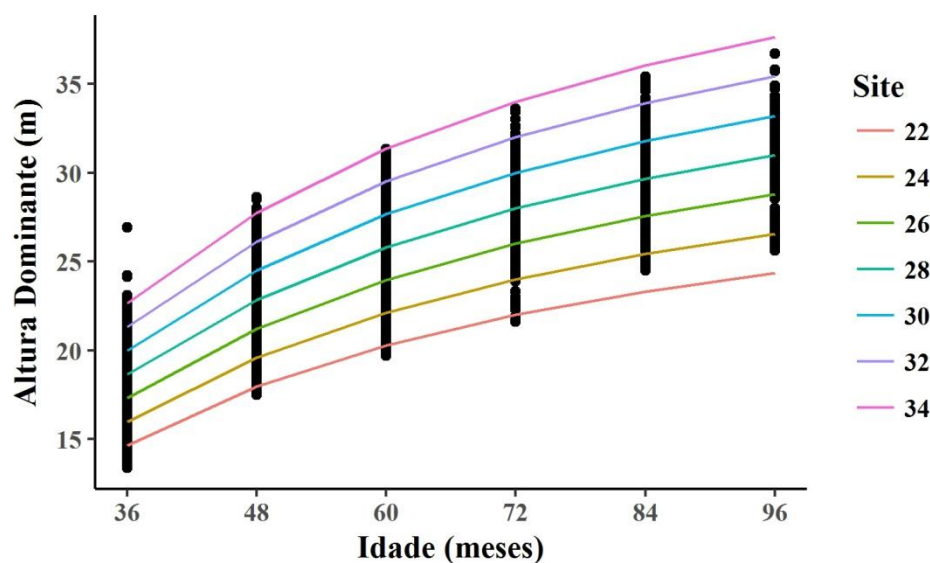


Figura 12 - Curvas de índices de local de *Eucalyptus* ssp. para uma idade índice de 72 meses pelo método da curva-guia.

Observa-se que as curvas de índice de local obtidas pelo método da curva-guia não abrangeram todos os valores observados de altura dominante (Figura 12). Analisando os dados de medição para alguns talhões que não foram abrangidos pelo feixe de curvas aos 36 meses de idade, percebeu-se que esses talhões apresentaram tendência de crescimento diferente do comportamento médio representado pelas curvas de índice de local (Figura 13).

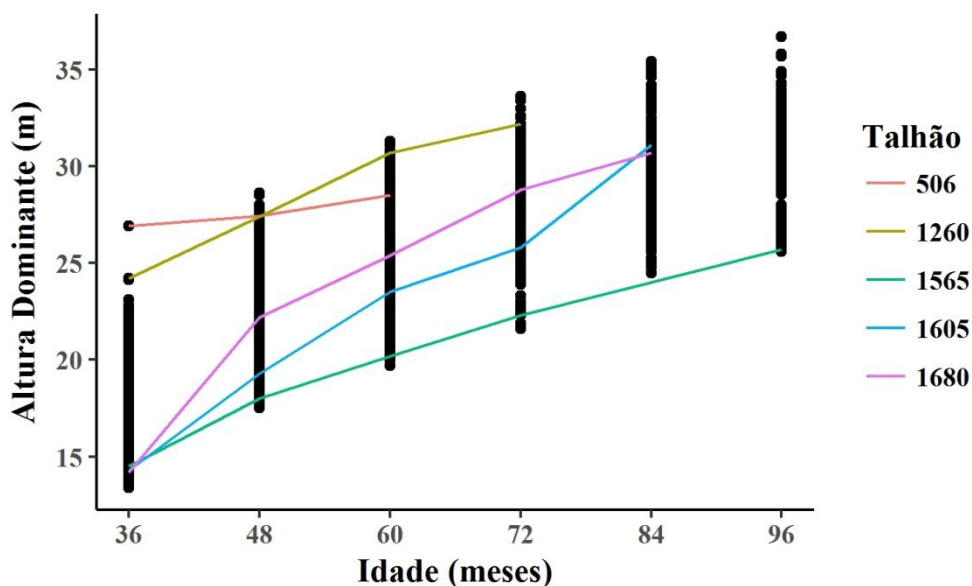


Figura 13 – Curvas de crescimento em altura dominante para alguns talhões que não foram abrangidos pelo feixe de curvas aos 36 meses de idade.

A Figura 13 mostra que há diferentes tendências de crescimento em altura dominante. Observa-se, por exemplo, que aparentemente o talhão 506 estagnou o crescimento em altura dominante a partir dos 36 meses de idade, enquanto o talhão 1260 apresenta tendência de estagnação aos 60 meses. Já o talhão 1605 apresenta inflexão na curva aos 72 meses de idade. Portanto, isso é um indicativo de que um método de construção de curvas de índice de local anamórficas, ou seja, pressupor uma mesma tendência de crescimento em altura dominante para os diferentes locais pode não ser adequado para este povoamento florestal.

Embora, nessa situação, esse tipo de método possa ser empregado ao estratificar o povoamento florestal. Há diversas possibilidades de estratificação para classificação da capacidade produtiva, sendo comum utilizar uma estratificação administrativa em que os estratos são definidos pelas diferentes regiões de plantio. No entanto, a estratificação baseada em características dos dados tem um potencial maior para ganhos na precisão de ajustes que a estratificação puramente administrativa (CAMARGOS, 2017).

Nesse sentido, uma alternativa eficiente é realizar a estratificação pela tendência de crescimento. Oliveira et al. (2008) agruparam clones com tendências de crescimento semelhantes para classificação da capacidade produtiva por meio do método da curva-guia, já Rabelo (2014) optou por agrupar talhões com tendências de crescimento semelhantes.

Ao estratificar em nível de talhão, como foi realizado neste estudo, deve-se ficar atento a quantidade de dados para ajuste, bem como se estes dados representam uma tendência definida de crescimento. Esse tipo de estratificação é importante porque em geral os ajustes para estimativas de altura requerem estratos menores, dessa forma é possível estimar o índice de local de forma mais precisa. Contudo, nesse tipo de ajuste o feixe de curvas obtido também apresentará anamorfismo, pois são utilizados parâmetros médios para o povoamento para construção das curvas de índice de local. Mas vale lembrar que o objetivo de construir curvas de índice de local é simplesmente ilustrar o padrão de crescimento em altura dominante. A etapa mais importante do processo é obter estimativas precisas de índice de local, pois essa variável é utilizada em modelos de crescimento e produção. Logo, para esse tipo de modelo, estimativas precisas de crescimento e produção são dependentes de estimativas precisas de índice de local (OLIVEIRA et al., 2008).

Na sequência das análises, estimou-se o índice de local utilizando os parâmetros obtidos para cada talhão e foram definidas três classes de qualidade de local, com amplitude de 4,8 m (Tabela 5).

Tabela 5 - Descrição das classes de capacidade produtiva superior (I), média (II) e inferior (III) obtidas pelo método da curva-guia, em que N é o número de talhões e os valores médios para as variáveis dendrométricas: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante e q = diâmetro quadrático médio

Classe	Amplitude	Idade (meses)	Área (ha)	<i>N</i>	<i>V</i> (m³.ha ⁻¹)	<i>B</i> (m².ha ⁻¹)	<i>Ht</i> (m)	<i>Hd</i> (m)	<i>q</i> (cm)
I	30,3-35,0	36	56,80	3	88,14	10,83	17,6	19,4	11,4
		48	223,28	9	172,40	14,86	22,3	25,7	13,2
		60	327,45	13	234,32	18,08	24,6	28,9	14,1
		72	641,13	24	346,62	21,03	28,8	31,6	16,0
		84	601,97	25	327,22	23,39	30,7	33,8	16,5
		96	178,25	11	322,72	22,31	29,9	34,7	16,4
Subtotal		2.028,88	85						
II	25,5-30,2	36	31,13	2	54,03	8,54	14,4	15,7	10,2
		48	3.352,04	125	132,72	13,29	20,5	22,5	12,2
		60	2.828,51	104	187,39	15,93	23,4	25,6	13,3
		72	1.544,23	67	232,51	18,41	25,6	27,7	14,3

Continua

Tabela 5 – Continuação

Classe	Amplitude	Idade (meses)	Área (ha)	<i>N</i>	<i>V</i> (m³.ha⁻¹)	<i>B</i> (m².ha⁻¹)	<i>Ht</i> (m)	<i>Hd</i> (m)	<i>q</i> (cm)
II	25,5-30,2	84	1.874,19	94	222,50	18,61	25,0	29,2	15,1
		96	2.344,19	119	252,09	19,71	27,2	31,1	15,4
	Subtotal		11.974,29	511					
III	20,7-25,4	48	1.543,60	54	104,91	11,92	19,0	20,5	11,7
		60	1.066,95	39	147,03	14,27	21,3	22,9	12,6
		72	142,36	7	186,53	15,98	23,0	25,1	13,5
		84	467,96	29	174,04	16,61	22,1	26,5	14,4
		96	540,10	25	189,87	17,95	23,1	26,6	14,9
	Subtotal		3.760,97	154					
Total		17.764,14	750						

As classes de produtividade superior, média e inferior representaram 11%, 68% e 21% do número de talhões (Tabela 5), respectivamente. De acordo com as classes obtidas pelo método da curva-guia, é possível sugerir que o potencial de produção apresenta característica de normalidade, com maior número de talhões com potencial de produção médio. Em termos de área, a proporção foi semelhante à encontrada para número de talhões com 12%, 67% e 21% para as classes superior, média e inferior, respectivamente.

Esse comportamento pode ser explicado pelo teorema do limite central (TCL). Esse teorema demonstra que: “se há um grande número de variáveis aleatórias independentes e com distribuição idêntica, então, com poucas exceções, a distribuição de suas somas tende à distribuição normal conforme o número dessas variáveis aumenta indefinidamente” (GUJARATI e PORTER, 2011 p. 119). Portanto, como há uma diversidade de fatores que atuam conjuntamente na capacidade produtiva do local, a soma de todos esses fatores tende à normalidade.

No entanto, os resultados obtidos podem ser influenciados pelo tipo de amostragem utilizada. Cabe lembrar que é comum empregar a amostragem aleatória nos inventários florestais contínuos. Esse tipo de amostragem resulta em alta frequência de parcelas na condição média dos povoamentos, o que nem sempre garante representatividade nas condições extremas (BEZERRA, 2009). Ao classificar a capacidade produtiva é importante garantir que todas as classes de produtividade estejam representadas na amostra de parcelas permanentes (CAMPOS e LEITE, 2013). Portanto, a situação ideal para a classificação da capacidade produtiva é utilizar a

amostragem seletiva, lançando parcelas de forma semelhante entre as diferentes classes, quando possível.

5.2 Método da predição dos parâmetros

Os parâmetros médios estimados para o modelo logístico em nível de parcela foram: $\hat{\beta}_0 = 33,259$; $\hat{\beta}_1 = 11,194$ e $\hat{\beta}_2 = 0,053$.

As equações obtidas ao relacionar os parâmetros do modelo logístico e o índice de site por parcela foram:

$$\beta_0 = 3,465 + 1,114*S \quad \bar{R}^2 = 6,36 \% \quad (24)$$

$$\beta_1 = -29,292 + 3,614^{ns}S - 0,079^{ns}S^2 \quad \bar{R}^2 = 0,02\% \quad (25)$$

$$\beta_2 = 0,133 - 0,006*S + 9,98e^{-5}*S^2 \quad \bar{R}^2 = 0,004\% \quad (26)$$

* significativo a 5% de significância pelo teste t , ^{ns} não significativo a 5% de significância pelo teste t .

Os baixos valores de \bar{R}^2 mostram que não há relação entre o índice de local e os parâmetros obtidos para o modelo logístico que possa ser modelada pelas equações ajustadas. O \bar{R}^2 avaliado separadamente não é uma estatística eficiente para confirmar esta relação, porém a falta de relação entre o índice de local e os parâmetros pode ser confirmada ao analisar a Figura 14.

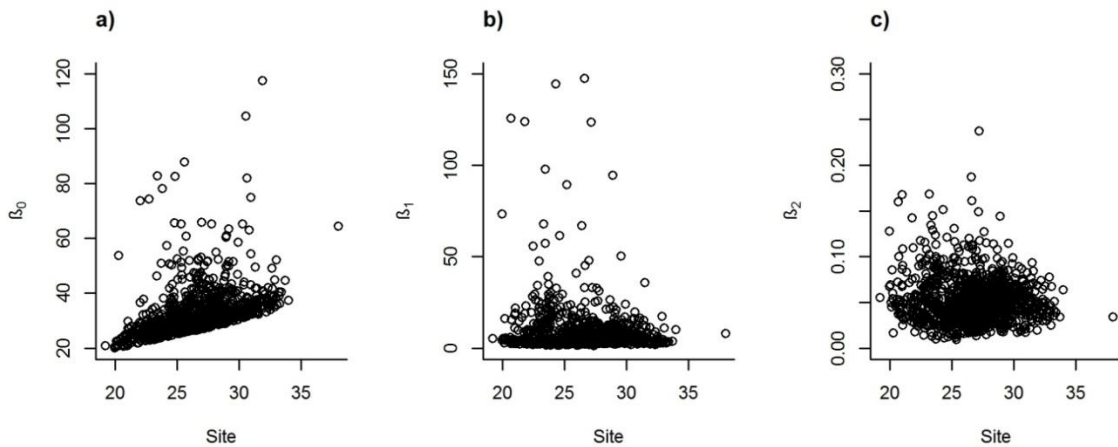


Figura 14 - Dispersão dos parâmetros estimados do modelo logístico em função dos índices de local por parcela. Em que: a) é a dispersão para o parâmetro $\hat{\beta}_0$, b) é a dispersão para o parâmetro $\hat{\beta}_1$ e c) é a dispersão para o parâmetro $\hat{\beta}_2$.

Como pode ser observado na Figura 14 não existe uma tendência bem definida entre os parâmetros e o índice de local. Dias et al. (2005) também não encontraram relação entre os parâmetros estimados para o modelo de Chapman-Richards e os índices de local estimados por parcela. Diante disso, esses autores utilizaram valores médios dos parâmetros ajustados em nível de parcela. No entanto, ao adotar esse procedimento não é possível construir curvas de índice de local polimórficas.

Nesse sentido, optou-se por testar o procedimento adotado por Taqueti (2014) ao utilizar o método da predição de parâmetros em modelos de produção, em que esses modelos foram ajustados por classe de índice de local e posteriormente os parâmetros obtidos foram relacionados com o centro de classe.

Dessa forma, a partir do índice de local estimado para cada parcela, os dados foram estratificados em quatro classes. O número de classes foi determinado visando obter representatividade das tendências. Posteriormente, o modelo logístico foi ajustado para cada classe e em seguida os parâmetros ajustados foram relacionados com o centro de classe de índice de local. A dispersão para os parâmetros é apresentada na Figura 15.

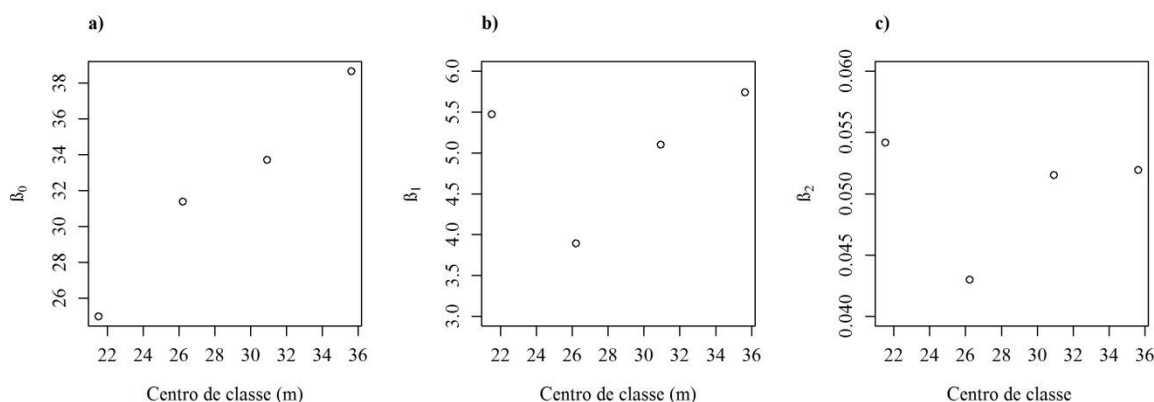


Figura 15 - Dispersão dos parâmetros estimados para o modelo logístico em função dos centros de classe de índice de local. Em que: a) é a dispersão para o parâmetro $\hat{\beta}_0$, b) é a dispersão para o parâmetro $\hat{\beta}_1$ e c) é a dispersão para o parâmetro $\hat{\beta}_2$.

As relações funcionais foram selecionadas por meio da ferramenta CurveFinder do *software* CurveExpert, conforme Taqueti (2014). Foi selecionado o modelo (27) para os parâmetros β_0 e β_2 ; e o modelo (28) para o parâmetro β_1 :

$$Y = \alpha_0 - \alpha_1 e^{-\alpha_2 x^{\alpha_3}} + \varepsilon \quad (27)$$

$$Y = \frac{\alpha_0 \alpha_1 + \alpha_2 x^{\alpha_3}}{\alpha_1 + x^{\alpha_3}} + \varepsilon \quad (28)$$

Substituindo o ajuste dos modelos (27) e (28) no modelo logístico, obteve-se a equação:

$$Hd = \frac{47,436 - 44,951e^{-0,005S^{1,607}}}{1 + \left(\frac{4,080 * 859,065 + 18,013S^{1,249}}{859,065 + S^{1,249}} \right) e^{-(0,060 - 0,025e^{-1,154S^{-0,065}})I}} \quad (29)$$

A equação (29) foi utilizada para estimar o índice de local por talhão e para construção de curvas de índice de local. O feixe de curvas obtido pelo método da predição dos parâmetros é apresentado na Figura 16.

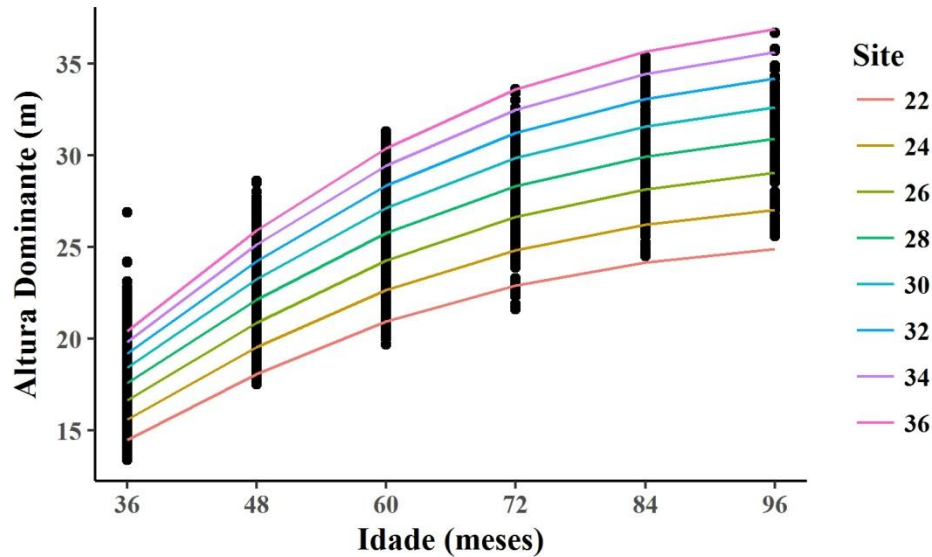


Figura 16 - Curvas de índices de local de *Eucalyptus* ssp. para uma idade índice de 72 meses pelo método da predição dos parâmetros.

Analisando a Figura 16 percebe-se que as curvas de índice de local apresentaram a mesma tendência, diferindo-se apenas na distância entre curvas. Também é possível observar a proximidade das curvas nas idades mais jovens. A menor distância entre curvas leva a uma classificação de sítios menos acurada em idades mais jovens, pois uma pequena imprecisão na medida de altura pode levar a uma classificação incorreta do sítio (MACHADO, 1980).

Os índices de local estimados foram agrupados em três classes de capacidade produtiva (Tabela 6), definidas com uma amplitude de 7,7 m.

Tabela 6 - Descrição das classes de capacidade produtiva superior (I), média (II) e inferior (III) obtidas pelo método da predição de parâmetros. Em que N é o número de talhões e os valores médios para as variáveis dendrométricas: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante e q = diâmetro quadrático médio

Classe	Amplitude	Idade (meses)	Área (ha)	N	V ($m^3.ha^{-1}$)	B ($m^2.ha^{-1}$)	Ht (m)	Hd (m)	q (cm)
I	36,1-43,7	48	111,21	4	172,95	14,91	22,9	27,0	13,2
		60	55,76	2	265,37	18,94	25,2	30,7	14,3
		Subtotal	166,97	6					
II	28,4-36,0	36	56,80	3	88,14	10,83	17,6	19,4	11,4
		48	1.884,04	65	144,85	13,83	21,3	23,6	12,5
		60	1.238,36	46	210,44	16,76	24,2	27,2	13,7
		72	1.103,30	43	309,64	20,28	27,9	30,5	15,4
		84	1.135,39	49	281,94	21,34	28,4	32,3	16,0
		96	1.342,12	72	282,54	20,82	28,6	32,5	15,8
		Subtotal	6.760,01	278					
III	20,7-28,3	36	31,13	2	54,03	8,54	14,4	15,7	10,2
		48	3.123,67	119	115,12	12,43	19,5	21,1	11,9
		60	2.928,79	108	167,20	15,18	22,4	24,3	13,0
		72	1.224,42	55	216,15	17,78	24,9	26,9	14,2
		84	1.808,73	99	205,33	17,88	23,9	28,0	14,9
		96	1.720,42	83	216,30	18,56	25,1	29,0	15,0
		Subtotal	10.837,16	466					
Total			17.764,14	750					

Observou-se que os índices de local estimados variaram de 20,7 m a 43,7 m (Tabela 6) enquanto a variação da altura dominante observada na idade índice foi de 21,6 a 33,6; então, aparentemente a altura dominante está sendo superestimada na idade índice para os talhões da classe I.

No entanto, ao analisar os dados percebeu-se que todos os talhões dessa classe não apresentaram medição aos 72 meses, sendo quatro talhões com medição até 48 meses e dois talhões com medição até os 60 meses. Dessa forma, não é possível afirmar que esses valores de altura não podem ser atingidos para o povoamento em estudo. Alguns talhões dessa classe

apresentaram incremento anual em altura de até 8m, de 36 para 48 meses. Para os talhões que apresentaram medição até os 60 meses, observou-se que até essa idade não havia tendência de estagnação, mas um incremento anual de 5 e 6 m, relativamente constante ao longo do tempo. Também é importante destacar que pelos métodos da curva-guia e equação das diferenças esses talhões foram alocados na classe I, mas com estimativas de índice de local inferiores às obtidas pelo método da predição dos parâmetros.

A maior proporção de talhões não ocorreu na classe II (Tabela 6) como verificado para o método da curva-guia. As classes de produtividade superior, média e inferior representaram 1%, 37% e 62% do número de talhões, respectivamente. No entanto, o teste de Shapiro-Wilk foi significativo a 10% de significância ($p = 9,802e^{-12}$) para as estimativas de índice de local, confirmando a hipótese de que o potencial de produção tende à normalidade.

O maior número de talhões ocorreu na classe III porque nela está contido o valor médio de índice de local (27,8 m). Dessa forma, é possível inferir que o cálculo da amplitude de classe (dado pela diferença entre o valor máximo e o valor mínimo dividido por três) nem sempre é adequado para classificar os talhões em produtividade superior, média e inferior. Nesse caso, por exemplo, se fossem adotados critérios como média \pm desvio-padrão para determinar os talhões pertencentes à classe II, e os valores superiores fossem alocados na classe I e inferiores na classe III, a frequência seria de 13%, 72% e 15% para as classes superior, média e inferior, respectivamente. Portanto, confirmando uma maior frequência na classe média.

5.3 Método da equação das diferenças

O valor médio para os parâmetros estimados para o método da equação das diferenças foi $\hat{\beta}_1 = -28,742$. O resumo dos ajustes é apresentado na Tabela 7.

Tabela 7 – Valores médios para os parâmetros estimados ($\hat{\beta}_1$), erro-padrão da regressão em porcentagem ($S_{yx\%}$) e raiz quadrada do erro médio (RQEM) para ajustes do modelo (17). Em que N é o número de ajustes

Nível	N	$\hat{\beta}_1$	$S_{yx\%}$	RQEM
Talhão	557	-29,394	4,78	1,31
Projeto/Clone	109	-26,212	5,25	1,15
Projeto	49	-25,878	5,69	1,25
Clone	27	-29,767	5,74	3,88
Povoamento	8	-31,883	6,59	17,43

Analisando a Tabela 7 é possível perceber a mesma tendência observada nos ajustes para o método da curva-guia, ou seja, a precisão aumenta com a redução do estrato. Os dados de ajuste possibilitaram obter uma equação específica para o talhão em 74% dos casos e em apenas 1% do número de talhões foi necessário utilizar uma equação em nível de povoamento.

As curvas de índice de local (Figura 17) foram construídas utilizando a seguinte equação:

$$\ln(Hd) = \ln(S) - 28,742(72^{-1} - I^{-1}). \quad (24)$$

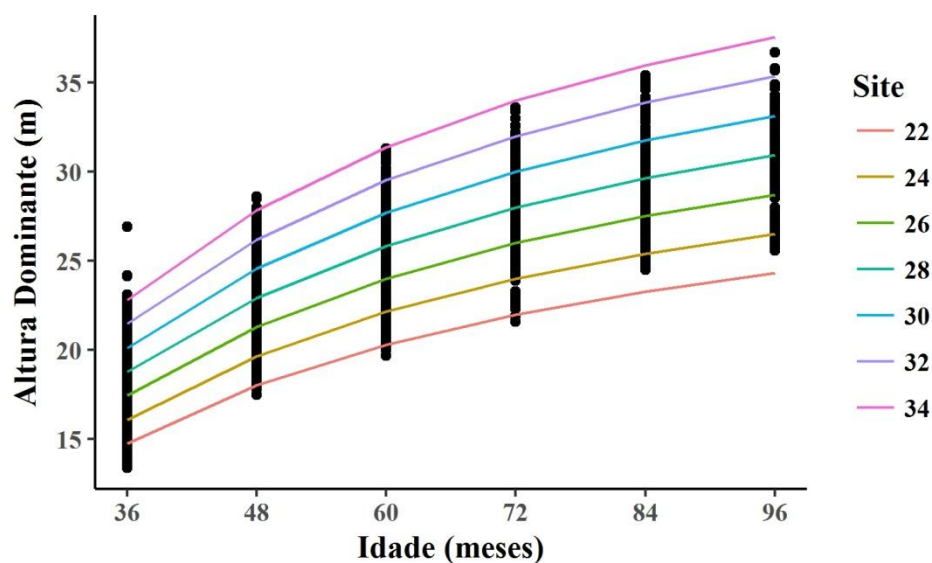


Figura 17 - Curvas de índices de local de *Eucalyptus spp.* para uma idade índice de 72 meses pelo método da equação das diferenças.

O feixe de curvas de índice de local obtido pelo método da equação das diferenças (Figura 17) foi muito semelhante às curvas obtidas pelo método da curva-guia (Figura 12). Assim como no método da curva-guia uma limitação são as diferentes tendências de crescimento o que pode ser minimizado com a estratificação do povoamento, conforme discutido na seção 5.1.

Os índices de local estimados foram agrupados em três classes de capacidade produtiva (Tabela 8), definidas com uma amplitude de 4,5 m. A proporção de talhões por classe de capacidade produtiva (Tabela 8) também foi muito semelhante aos valores encontrados para o método da curva-guia, com 13%, 66% e 21% para as classes superior, média e inferior.

Tabela 8 - Descrição das classes de capacidade produtiva superior (I), média (II) e inferior (III) obtidas pelo método da equação das diferenças. Em que N é o número de talhões e os valores médios para as variáveis dendrométricas: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante e q = diâmetro quadrático médio

Classe	Amplitude	Idade (meses)	Área (ha)	<i>N</i>	<i>V</i> (m³.ha ⁻¹)	<i>B</i> (m².ha ⁻¹)	<i>Ht</i> (m)	<i>Hd</i> (m)	<i>q</i> (cm)	
I	29,9-34,4	36	36,50	2	88,39	10,68	17,6	19,6	11,4	
		48	415,81	15	163,68	14,26	22,1	25,2	13,0	
		60	531,36	20	229,04	17,78	24,4	28,3	14,0	
		72	670,39	26	341,10	20,97	28,7	31,5	16,0	
		84	601,97	25	327,22	23,39	30,7	33,8	16,5	
		96	179,07	11	321,11	22,30	29,7	34,7	16,4	
Subtotal			2.435,10	99						
II	25,4-29,8	36	32,85	2	73,25	10,02	16,3	17,6	10,8	
		48	2.957,44	109	133,28	13,37	20,5	22,5	12,2	
		60	2.700,41	99	184,10	15,80	23,3	25,4	13,3	
		72	1.525,24	66	230,38	18,33	25,5	27,6	14,3	
		84	1.983,68	100	219,69	18,49	24,8	29,0	15,1	
		96	2.343,37	119	252,24	19,71	27,2	31,1	15,4	
Subtotal			11.542,99	495						
III	20,8-25,3	36	18,58	1	49,20	8,16	13,9	15,1	10,1	
		48	1.745,67	64	106,63	11,98	19,2	20,6	11,8	
		60	991,14	37	147,61	14,29	21,3	22,9	12,6	
		72	132,09	6	188,24	15,82	23,1	25,1	13,4	
		84	358,47	23	173,61	16,64	22,0	26,3	14,4	
		96	540,10	25	189,87	17,95	23,1	26,6	14,9	
Subtotal			3.786,05	156						
Total				17.764,14	750					

5.3 Teste de anamorfismo

Os resultados para o primeiro teste de anamorfismo são apresentados na Tabela 9.

Tabela 9 - Coeficiente de variação para a variável altura dominante por classe de idade e índice de local obtidos pelos métodos da curva-guia, predição dos parâmetros e equação das diferenças

Método	Classe	Idade (meses)					
		36	48	60	72	84	96
Curva-guia	I	2,6	5,3	3,9	3,0	2,6	3,0
	II	5,6	5,0	4,1	4,2	4,2	4,1

Tabela 9 – Continuação

Método	Classe	Idade (meses)					
		36	48	60	72	84	96
Curva-guia	III	-	5,6	4,5	1,3	3,1	2,9
	I	-	2,6	0,5	-	-	-
Predição dos parâmetros	II	2,6	3,5	3,6	4,8	5,2	3,5
	III	5,6	4,8	5,4	3,8	4,6	5,9
Equação das diferenças	I	1,6	5,2	4,4	3,3	2,6	2,9
	II	10,3	4,7	4,1	4,1	4,4	4,1
	III	-	5,5	4,6	1,3	3,0	2,9

De acordo com os resultados apresentados na Tabela 9 é possível inferir que a altura média das árvores dominantes apresenta comportamento polimórfico. Segundo Scolforo (2006), as curvas anamórficas devem apresentar estabilidade no coeficiente de variação por classe de índice de local e idade. Esse comportamento não foi observado para as classes de capacidade produtiva para os três métodos analisados. No entanto, uma limitação do teste é que não há referência do quão semelhante deve ser o coeficiente de variação para que as curvas sejam classificadas em anamórficas. Além disso, o tamanho da amostra pode causar distorções no cálculo do coeficiente de variação (SCOLFORO, 2006), pois o número de talhões por classe de índice de local não é constante, comprometendo assim os resultados.

Os resultados do segundo teste são apresentados na Tabela 10.

Tabela 10 - Parâmetros estimados ($\hat{\beta}_0$ e $\hat{\beta}_1$) para um modelo linear simples (modelo 20) relacionando índice de local e altura dominante por classe de idade para os métodos da curva-guia, predição dos parâmetros e equação das diferenças

Método	Parâmetro	Idade (meses)					
		36	48	60	72	84	96
Curva-guia	$\hat{\beta}_0$	0,000	1,317	-3,153	0,000	-0,218	0,347
	$\hat{\beta}_1$	1,761	1,150	1,209	1,000	0,947	0,888
Predição dos parâmetros	$\hat{\beta}_0$	-11,583	-12,296	-8,208	-9,344	-7,711	-5,846
	$\hat{\beta}_1$	2,280	1,832	1,412	1,325	1,199	1,099
Equação das diferenças	$\hat{\beta}_0$	0,000	-0,813	-2,833	0,000	-0,109	0,341
	$\hat{\beta}_1$	1,557	1,242	1,196	1,000	0,945	0,891

Analisando a Tabela 10 é possível observar que a interseção é igual a 0 e a inclinação igual a um na idade índice para os métodos da curva-guia e equação dos parâmetros. Além disso,

para esses métodos, a inclinação é menor que 1 em idades superiores à idade índice e maior que 1 nas idades inferiores à idade índice, sendo maior quanto mais jovem a idade considerada. De acordo com Scolforo (2006) isso indica comportamento anamórfico para as curvas de índices de local obtidas para os métodos da curva-guia e equação das diferenças. Já o método da predição dos parâmetros não apresentou essas características, indicando que as curvas obtidas por esse método são polimórficas.

Portanto, os testes aplicados levaram a diferentes resultados não permitindo concluir se o padrão de crescimento em altura dominante é anamórfico ou polimórfico. Além disso, pode não ser adequado testar o anamorfismo ou polimorfismo a partir dos índices de local estimados, pois o padrão de crescimento em altura dominante pode ser imposto pelo método de construção de curvas de índice de local empregado. Diante disso, foi aplicado um teste de identidade de modelos (REGAZZI, 1993) utilizando dados de altura dominante para verificar se a tendência de crescimento é semelhante para as diferentes classes de local.

Para determinar as classes para aplicar o teste foram utilizados valores de altura dominante na idade índice. Para os talhões que não apresentaram medição aos 72 meses, a altura dominante foi estimada utilizando os parâmetros obtidos em ajustes para o método da curva-guia.

A partir desses valores de altura dominante o povoamento foi estratificado em três classes de capacidade produtiva. Foi utilizado o modelo de Schumacher (1939) linearizado para descrever a tendência de crescimento em altura dominante. Procedeu-se à verificação da igualdade entre as curvas obtidas para as três classes de capacidade produtiva por meio de um teste de identidade de modelos lineares, conforme Regazzi (1993).

Testou-se a igualdade para os parâmetros β_0 e β_1 (p-valor = $1.537943e^{-245}$) e apenas para o parâmetro β_1 (p-valor = $1.196025e^{-101}$). Em ambas as situações o teste foi significativo a 5% de significância, permitindo concluir que há diferença na tendência de crescimento em altura dominante por classe de local, ou seja, o padrão de crescimento é polimórfico.

5.4 Redes neurais artificiais

A análise discriminante para os testes de 1 a 5 é apresentada na Tabela 11.

Tabela 11 – Número de talhões e porcentagem de acerto segundo a análise discriminante (AD) para a classificação da capacidade produtiva por meio de rede neural artificial (RNA) utilizando diferentes entradas. Em que: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante, q = diâmetro quadrático médio

	Classe	I	II	III	RNA	Acerto (%)
V	I	26	25	3	54	48,1
	II	6	346	86	438	79,0
	III	0	166	92	258	35,7
	AD	32	537	181	750	61,9
B	I	10	13	3	26	38,5
	II	1	201	100	302	66,6
	III	0	86	336	422	79,6
	AD	11	300	439	750	72,9
Ht	I	1	19	1	21	4,8
	II	0	153	152	305	50,2
	III	0	94	330	424	77,8
	AD	1	266	483	750	64,5
Hd	I	0	14	11	25	0,0
	II	0	18	154	172	10,5
	III	0	7	546	553	98,7
	AD	0	39	711	750	75,2
q	I	0	38	12	50	0,0
	II	0	648	0	648	100,0
	III	0	45	7	52	13,5
	AD	0	731	19	750	87,3

No teste 6 utilizou-se a análise de componentes principais para determinar as variáveis mais importantes para entrada na rede. Os resultados obtidos são apresentados na Tabela 12.

Tabela 12 – Autovetores, proporção da variância e variância acumulada para os componentes principais analisados

Componente Principal	Autovetor						Proporção da Variância (%)	Variância Acumulada (%)
	V	B	Ht	Hd	q	Arv_ha		
Y_1	0,457	0,442	0,487	0,469	0,371	0,045	59,7	59,7
Y_2	0,220	0,230	-0,035	-0,126	-0,439	0,830	20,8	80,5

Tabela 12 – Continuação

Componente Principal	Autovetor						Proporção da Variância (%)	Variância Acumulada
	<i>V</i>	<i>B</i>	<i>Ht</i>	<i>Hd</i>	<i>q</i>	<i>Arv_ha</i>		
Y_3	-0,042	-0,599	0,432	0,505	-0,443	0,037	8,0	88,5
Y_4	0,549	0,191	-0,016	-0,186	-0,583	-0,536	7,6	96,0
Y_5	-0,554	0,567	-0,150	0,454	-0,354	-0,134	2,2	98,3
Y_6	0,364	-0,188	-0,743	0,519	0,089	0,050	1,8	100,0

Analisando a Tabela 12 percebe-se que o primeiro componente principal foi responsável por 59,7% da variância dos dados. Observa-se também que a variável *Ht* foi a de menor importância para explicar a variação dos dados, pois apresentou maior valor absoluto de autovetor para o último componente (Y_6). Em outras palavras, foi a variável de maior peso no componente de menor contribuição para a variância total dos dados. Assim, a classificação foi realizada novamente retirando a variável *Ht*.

Comparando a classificação sem a variável *Ht* com a classificação com as variáveis *V*, *B*, *Hd*, *q* e *Arv_ha* observou-se que 170 talhões mudaram de classe. Além disso, a porcentagem de acerto geral da análise discriminante reduziu de 83,6% para 78,9%. Nesse sentido, a variável *Ht* foi mantida e o processo encerrado.

Portanto, embora a variável *Ht* tenha sido a menos importante de acordo com a análise de componentes principais, a retirada dessa variável alterou os resultados mostrando que a *Ht* foi importante para obtenção das classes de capacidade produtiva. Assim, é necessário cautela ao utilizar a análise de componentes principais, pois apenas a contribuição na variação total dos dados pode não ser um critério adequado para seleção de variáveis. Uma justificativa para isso é que se trata de dados biológicos, e nesses casos a tomada de decisão não deve ser meramente estatística. Essa mesma situação ocorre, por exemplo, em ajustes de modelos de regressão em que uma variável independente pode não ser significativa estatisticamente, mas biologicamente explicar a variável dependente, justificando mantê-la no modelo.

A análise discriminante para o teste 6 é apresentada na Tabela 13.

Tabela 13 – Número de talhões e porcentagem de acerto segundo a análise discriminante (AD) para a classificação da capacidade produtiva por meio de rede neural artificial (RNA). Em que: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante, q = diâmetro quadrático médio e Arv_ha = número de árvores por hectare

Entrada	Classe	I	II	III	RNA	Acerto (%)
V, B, Ht, Hd, q, Arv_ha	I	32	36	1	69	46,4
	II	15	576	19	610	94,4
	III	0	52	19	71	26,8
	AD	47	664	39	750	83,6

Analisando as Tabelas 11 e 13, observa-se que a maior porcentagem de acerto geral foi para a variável q , com 87,3%. No entanto, analisando a porcentagem de acerto por classe de capacidade produtiva para essa variável, enquanto a classe II apresentou 100% de acerto, a classe I apresentou 0%. Isso aconteceu porque pela análise discriminante nenhum talhão apresentou maior probabilidade de ocorrência na classe I, ou seja, foram obtidas apenas duas classes de capacidade produtiva pela análise discriminante.

A distribuição da variável q na rede foi relativamente homogênea (Figura 18). A maioria dos talhões apresentou q até 15 cm, sendo que a média para essa variável aos 72 meses foi 14,7 cm. Isso pode ser explicado pela distribuição diamétrica típica de florestas equiâneas, que tende à normalidade, isto é, com maior frequência próximo ao ponto médio. Dessa forma, utilizar apenas essa variável pode não ser suficiente para discriminar os talhões em três classes de capacidade produtiva. O mesmo comportamento foi observado para a variável Hd (Tabela 11), possivelmente pela mesma razão identificada para a variável q .

A segunda maior porcentagem de acerto geral foi para o teste 6 (Tabela 13) com 83,6%; sendo possível obter três classes de capacidade produtiva segundo a análise discriminante. Por essa razão, a rede neural treinada com o conjunto de variáveis do teste 6 foi selecionada para comparação com os outros métodos de classificação. A descrição das classes de capacidade produtiva obtidas a partir da entrada selecionada é apresentada na Tabela 14.

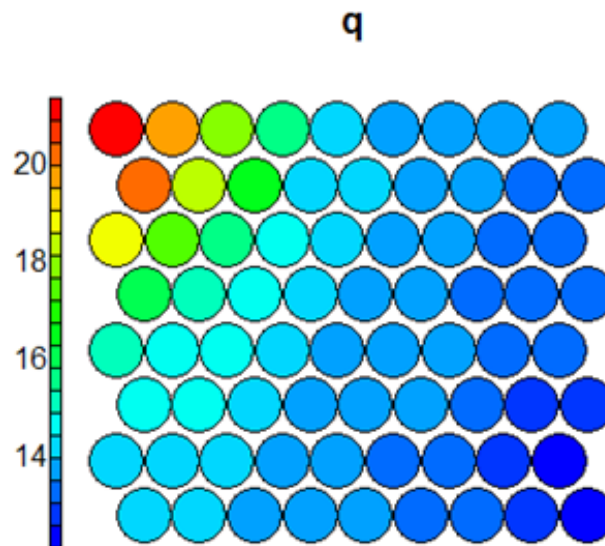


Figura 18 – Plano de componentes para a rede neural artificial treinada com a variável diâmetro quadrático médio (q).

Tabela 14 - Descrição das classes de capacidade produtiva superior (I), média (II) e inferior (III) obtidas por meio de rede neural artificial. Em que N é o número de talhões e os valores médios para as variáveis dendrométricas: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante e q = diâmetro quadrático médio

Classe	Idade (meses)	Área (ha)	N	V ($m^3 \cdot ha^{-1}$)	B ($m^2 \cdot ha^{-1}$)	Ht (m)	Hd (m)	q (cm)
I	36	56,80	3	88,14	10,83	17,6	19,4	11,4
	48	240,21	9	187,06	16,86	22,3	25,4	13,3
	60	218,45	8	269,54	19,07	25,9	28,8	14,4
	72	414,54	18	373,02	21,76	29,4	31,5	16,1
	84	598,71	27	314,97	22,96	29,8	32,7	16,3
	96	51,22	4	307,86	22,19	29,4	32,2	16,1
Subtotal		1.579,93	69					
II	36	31,13	2	54,03	8,54	14,4	15,7	10,2
	48	4.489,51	163	127,30	13,04	20,2	22,0	12,2
	60	3.885,69	140	179,82	15,74	23,0	25,2	13,2
	72	1.866,79	78	232,77	18,33	25,6	27,9	14,4
	84	2.019,36	99	222,70	18,77	25,0	29,0	15,2
	96	2.532,60	128	257,12	20,05	27,2	31,1	15,5
Subtotal		14.825,08	610					
III	48	389,20	16	85,87	10,03	18,5	20,2	11,2
	60	118,77	8	117,27	11,67	19,9	21,9	12,0
	72	46,39	2	166,46	14,45	22,1	24,9	13,4

Continua

Tabela 14 – Continuação

Classe	Idade (meses)	Área (ha)	<i>N</i>	<i>V</i> (m ³ .ha ⁻¹)	<i>B</i> (m ² .ha ⁻¹)	<i>Ht</i> (m)	<i>Hd</i> (m)	<i>q</i> (cm)
III	84	326,05	22	163,23	15,37	21,7	27,3	14,2
	96	478,72	23	180,55	16,71	23,4	27,7	14,5
Subtotal		1.359,13	71					
Total		17.764,14	750					

Na classificação por meio de rede neural artificial 9% dos talhões foram alocados na classe I, 81% na classe II e 10% na classe III. Por meio dos planos de componentes é possível fazer inferências da contribuição de cada variável na formação das classes (Figura 19). Esse tipo de visualização mostra a distribuição de cada variável individualmente na rede treinada. As variáveis são visualizadas indiretamente, por meio dos valores de cada componente dos vetores de pesos sinápticos que são utilizados para colorir o mapa.

Relacionando a Figura 19 com as classes obtidas, observa-se na Figura 20 que a região central da rede foi classificada em produtividade média, o canto inferior esquerdo da rede foi classificado em produtividade superior e canto inferior direito da rede foi classificado em produtividade inferior. Esse mesmo comportamento foi observado para a distribuição das variáveis *V*, *B*, *Ht* e *Hd* individualmente na rede (Figura 19), o que permite inferir que essas variáveis foram as que mais contribuíram para o agrupamento.

A variável *q* é a variável que apresenta distribuição mais homogênea na rede (Figura 19). Os valores de diâmetro da classe II são semelhantes à classe III (Figura 20), a discriminação das classes é evidente apenas para a classe I, o que justifica manter essa variável na análise.

O número de árvores por hectare na rede não foi determinante para a formação das classes. Apenas a classe I foi mais homogênea e a classe II foi a mais heterogênea quanto à densidade do povoamento.

É importante destacar que os planos de componentes também permitem visualizar as correlações entre as variáveis de entrada da rede (LEITE et al., 2006). Embora não seja possível quantificar o grau de similaridade entre as variáveis é possível afirmar, a partir da análise visual das relações existentes, que a correlação entre o número de árvores por hectare e cada uma das demais variáveis de povoamento é baixa. Isso é importante porque em geral as variáveis *V*, *B* e *q* não são utilizadas para classificar o potencial de produção porque são influenciadas pela densidade do povoamento.

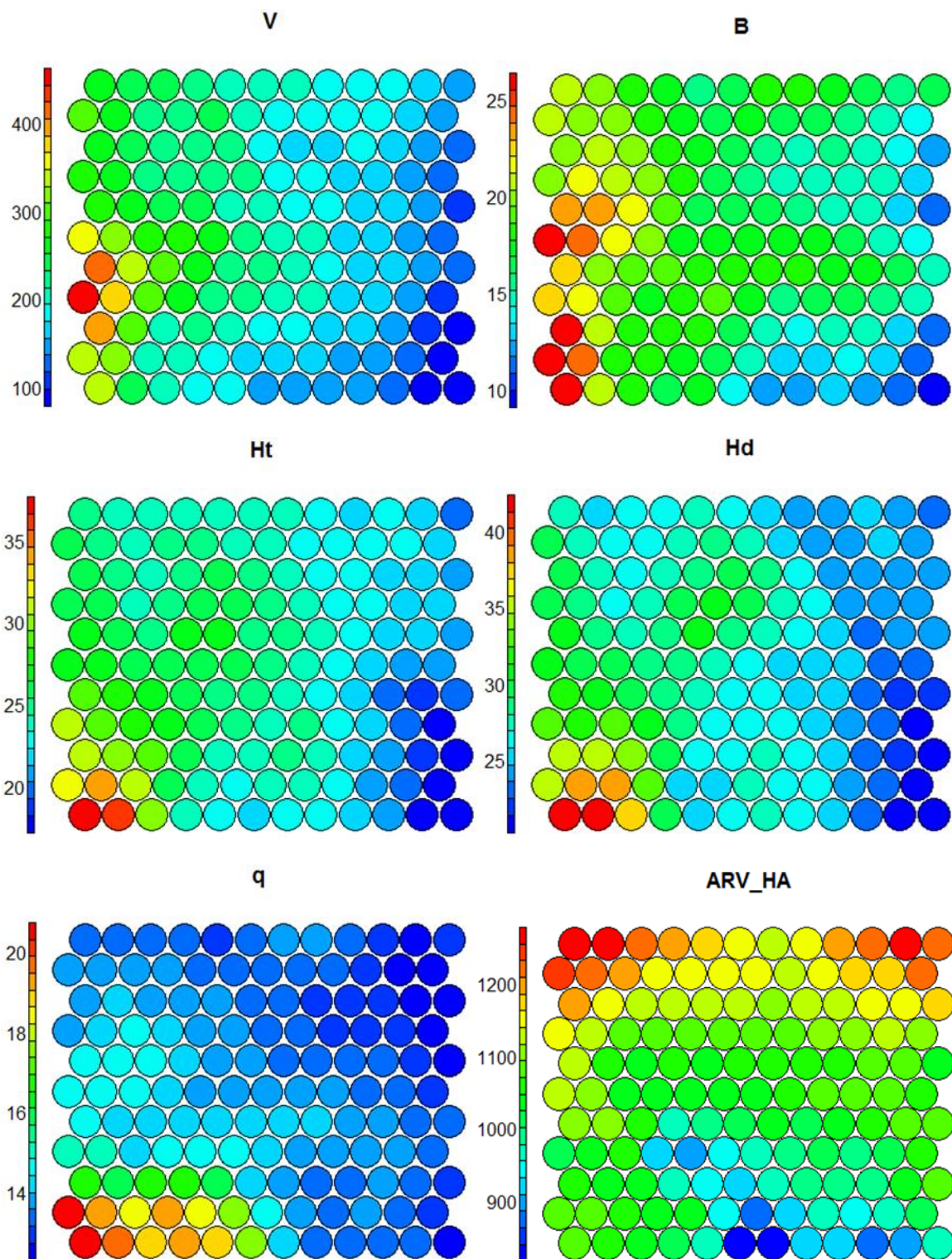


Figura 19 – Plano de componentes para as variáveis volume total com casca (V), área basal (B), altura total (Ht), altura dominante (Hd), diâmetro quadrático médio (q) e número de árvores por hectare (ARV_{HA}).

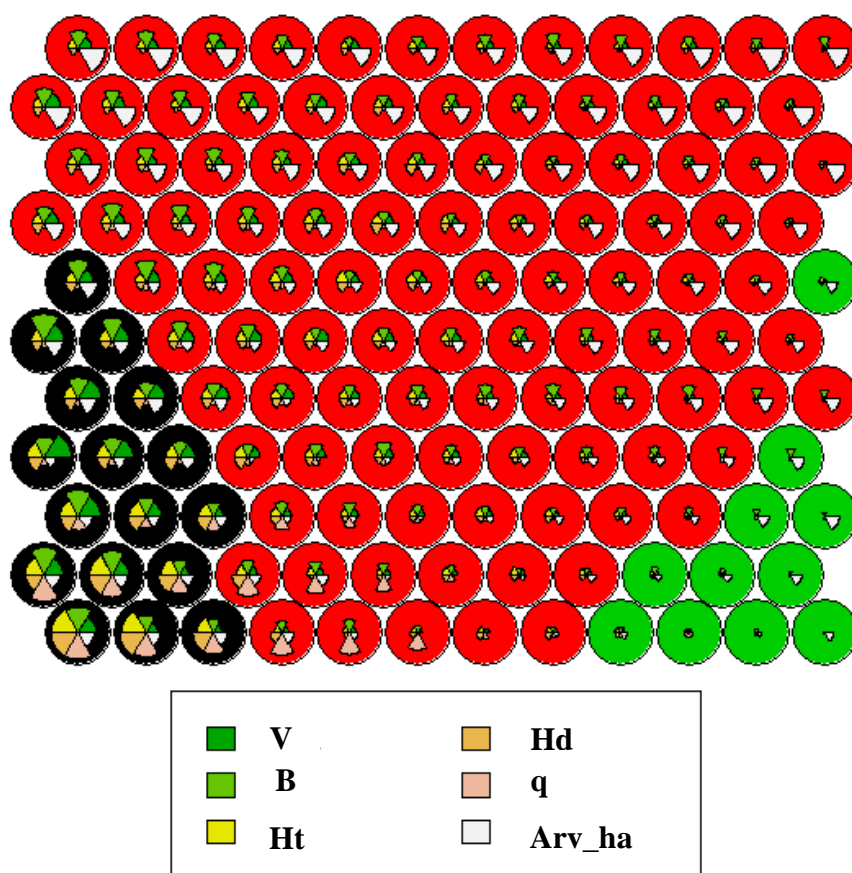


Figura 20 – Distribuição das variáveis nas classes de capacidade produtiva superior (preto), média (vermelho) e inferior (verde). Em que: V = volume total com casca, B = área basal, Ht = altura total, Hd = altura dominante, q = diâmetro quadrático médio e Arv_ha = número de árvores por hectare.

No entanto, observou-se que a distribuição dessas variáveis na rede não teve relação com a distribuição da variável número de árvores por hectare. Na classe II, por exemplo, enquanto há uma grande variação na densidade dentro dessa classe, a distribuição das outras variáveis é relativamente homogênea, sobretudo para o diâmetro quadrático médio. Uma possível explicação para isso, é que talvez até a idade de 72 meses a densidade de povoamento pode não ter influência significativa para essas variáveis. Diante disso, essa variável poderia ser retirada da análise, no entanto, optou-se por mantê-la.

É importante destacar que a visualização dos grupos é uma vantagem do agrupamento utilizando rede neural artificial em relação a outros métodos de agrupamento. A partir da visualização é possível fazer inferências de como os dados estão distribuídos, uma vez que se espera que as características estatísticas dos dados sejam preservadas (BRAGA et al., 2007).

Assim, uma das propriedades da rede auto-organizável é o casamento por densidade. De acordo com essa propriedade o mapeamento efetuado reflete a distribuição de probabilidade dos dados no espaço de entrada original. Dessa forma, mais neurônios serão utilizados para representar regiões onde os vetores de amostra são retirados com alta probabilidade de ocorrência, enquanto menos neurônios serão utilizados para representar regiões do espaço de entrada de onde vetores de amostra são retirados com uma baixa probabilidade de ocorrência (HAYKIN, 2001; HIGA et al., 2010).

Portanto, apenas com a análise visual é possível fazer inferências da distribuição de probabilidade do espaço de entrada. Para os dados utilizados, é possível identificar, por exemplo, que a probabilidade de ocorrência de talhões de alta produtividade é baixa, pois há poucos neurônios representando esse conjunto de amostras.

5.5 Comparação dos métodos

Para comparar a classificação pelos diferentes métodos foi calculada a porcentagem de coincidência na alocação dos talhões por classe de capacidade produtiva (Tabela 15).

Tabela 15 – Número de talhões por classe de capacidade produtiva e porcentagem de coincidência (C) entre os métodos empregados, em que RNA = rede neural artificial

Comparação	Classe	I	II	III	Total	C (%)
Curva-guia x Predição dos parâmetros	I	6	79	0	85	7,1
	II	0	196	315	511	38,4
	III	0	3	151	154	98,1
	Total	6	278	466	750	47,1
Curva-guia x Equação das diferenças	I	82	3	0	85	96,5
	II	17	482	12	511	94,3
	III	0	10	144	154	93,5
	Total	99	495	156	750	94,4
Curva-guia x RNA	I	51	34	0	85	60,0
	II	18	471	22	511	92,2
	III	0	105	49	154	31,8
	Total	69	610	71	750	76,1

Continua

Tabela 15 – Continuação

Comparação	Classe	I	II	III	Total	C (%)
Predição dos parâmetros x Equação das diferenças	I	6	0	0	6	100,0
	II	93	182	3	278	65,5
	III	0	313	153	466	32,8
	Total	99	495	156	750	45,5
Predição dos parâmetros x RNA	I	6	0	0	6	100,0
	II	56	220	2	278	79,1
	III	7	390	69	466	14,8
	Total	69	610	71	750	39,3
Equação das diferenças x RNA	I	52	47	0	99	52,5
	II	17	455	23	495	91,9
	III	0	108	48	156	30,8
	Total	69	610	71	750	74,0

Analisando a Tabela 15 percebe-se que a porcentagem geral de coincidência foi maior para os métodos da curva-guia e equação das diferenças (94,4%), seguida por curva-guia e RNA (76,1%). Uma possível explicação para a semelhança entre os métodos da curva-guia e equação das diferenças é que ambos produzem curvas anamórficas. Além disso, em ambos os métodos foi utilizado o modelo de Schumacher (1939) linearizado, sendo que os valores médios para o parâmetro $\hat{\beta}_1$ foram muito semelhantes (-29,236 para o método da curva-guia e -28,742 para o método da equação das diferenças).

O menor valor de porcentagem geral foi encontrado ao comparar o método da predição dos parâmetros e RNA (39,3%), isso porque 390 talhões foram alocados na classe II pela RNA e na classe III pelo método da predição dos parâmetros. Mas como discutido na seção 5.2, o valor médio de índice de local (27,8 m) obtido pelo método da predição dos parâmetros está contido na classe III. Isso também pode justificar o fato de 7 talhões da classe I obtida por RNA terem sido alocados na classe III do método da predição dos parâmetros. Dentre todas as comparações, essa foi a única situação de alternância da classe I para a classe III, evidenciando que de forma geral houve coerência para os resultados obtidos pelos diferentes métodos.

As divergências frequentes nas classificações ocorreram devido a alocação de talhões da classe I para a II e vice-versa ou da classe II para a classe III e vice-versa. Isso pode ocorrer, por exemplo, quando um talhão apresenta índice de local próximo a fronteiras de classe. Tendo em

vista que as amplitudes de classe são diferentes para cada método, pode levar à alocação em diferentes classes nesse caso.

Como para a RNA não há amplitude de classe, esse tipo de situação pode ser exemplificada por um talhão que apresenta capacidade produtiva máxima para a classe média (valores máximos para as variáveis dendrométricas na classe II), por exemplo, o que torna possível que este talhão também seja classificado na classe I por outro método ou no treinamento repetido da RNA, já que seu potencial produtivo é próximo do limite inferior da classe I.

Na Tabela 16, são apresentados os valores de área e volume total com casca por classe de capacidade produtiva para os quatro métodos empregados.

Tabela 16 – Área e volume total com casca (V) para as classes de capacidade produtiva superior (I), média (II) e inferior (III) obtidas pelos métodos da curva-guia, predição dos parâmetros, equação das diferenças e redes neurais artificiais (RNA)

Método	Classe	Área (ha)	V (m ³)
Curva-guia	I	2028,88	588.696,83
	II	11.974,29	2.344.287,91
	III	3.760,97	530.317,26
Predição dos parâmetros	I	166,97	33.567,81
	II	6.760,01	1.571.239,95
	III	10.837,16	1.858.494,24
Equação das diferenças	I	2435,1	669.528,55
	II	11.542,99	2.269.570,64
	III	3.786,05	524.202,80
RNA	I	1.579,93	471.269,93
	II	14.825,08	2.796.216,51
	III	1.359,13	195.815,56

Analisando a Tabela 16 é possível notar que os métodos da curva-guia e da equação das diferenças também foram muito semelhantes em termos de área e volume total. A classe II obtida por RNA apresentou maior área e volume, quando comparada à classe II dos outros métodos. Possivelmente isso ocorreu porque talhões que foram alocados na classe III pelos outros métodos foram alocados na classe II da classificação por RNA. O que pode ser confirmado ao verificar que a classe III da RNA apresenta menor área e volume que os outros métodos.

Com exceção do método da predição dos parâmetros, a tendência foi de maior estoque volumétrico e área para a classe de produtividade média.

Também foram calculadas diferenças para a variável volume total com casca em relação ao método da curva-guia, considerado padrão (Tabela 17) por ser o mais utilizado no Brasil.

Tabela 17 – Diferença em relação ao método da curva-guia para a variável volume total com casca por classe de capacidade produtiva – superior (I), média (II) e inferior (III) – obtidas pelos métodos da predição dos parâmetros, equação das diferenças e redes neurais artificiais (RNA)

Método	Classe	Diferença (m ³)	Diferença (%)
Predição dos parâmetros	I	-555.129,03	-94,3
	II	-773.047,96	-33,0
	III	1.328.176,99	250,4
Equação das diferenças	I	80.831,72	13,7
	II	-74.717,26	-3,2
	III	-6.114,46	-1,2
RNA	I	-117.426,90	-19,9
	II	451.928,60	19,3
	III	-334.501,70	-63,1

Conforme a Tabela 17, o método da equação das diferenças foi o mais semelhante à curva-guia em termos de estoque volumétrico. A semelhança entre as classes desses dois métodos também foi verificada para a coincidência na alocação de talhões por classe de capacidade produtiva e em termos de área. Isso ocorreu porque as estimativas de índice de local foram muito semelhantes para esses métodos, com isso as amplitudes de classe encontradas foram também muito próximas (4,5 e 4,8 m) levando a obter classes semelhantes de índice de local.

Já o método da predição dos parâmetros apresentou diferença percentual de até 250% em relação ao método da curva-guia. A diferença entre as classes desses métodos também foi evidenciada ao compará-los em termos de coincidência na alocação de talhões por classe de capacidade produtiva e em termos de área. As estimativas de índice de local obtidas pelo método da predição dos parâmetros não foram próximas das estimativas de índice de local pelo método da curva-guia.

A maior diferença percentual de estoque volumétrico da classificação por RNA em relação à curva-guia foi de 63,1% para a classe III, porque como observado nas comparações

anteriores 105 talhões classificados na classe II pela RNA, foram alocados na classe III pelo método da curva-guia.

Diante dos resultados obtidos é possível inferir que a classificação por meio de rede neural foi eficiente. O uso dessa técnica pode ser promissor por possibilitar a inclusão de outras variáveis associadas à qualidade do local como, por exemplo, tipo de solo, precipitação, dentre outras. Essa abordagem alternativa para a classificação do potencial produtivo é interessante porque ao incluir mais variáveis no processo espera-se que a classificação seja mais eficiente que utilizar apenas uma variável.

Além disso, a classificação por meio de RNA não tem restrição quanto ao padrão de crescimento como a curva-guia, em que deve ser utilizada uma equação específica para cada local em povoamentos que apresentam tendência de crescimento polimórfica em altura dominante.

No entanto, a maior vantagem de se utilizar o método da curva-guia é a praticidade na aplicação, assim como o método da equação das diferenças. Já o método da predição dos parâmetros pode não ser muito prático, quando utilizado um modelo não linear, devido à dificuldade de obtenção de parâmetros iniciais que possibilitem a convergência do algoritmo de ajuste.

Um inconveniente da classificação por meio de rede neural artificial é a condição de que os talhões devem apresentar a mesma idade, o que torna necessário realizar estimativas para as variáveis de entrada, na idade de referência. Embora essa dificuldade seja dependente do programa utilizado para processamento dos dados. Pois, ao utilizar um programa como o R, simplifica-se e agiliza-se o processamento. Soma-se a isso a possibilidade de utilizar redes neurais artificiais nessa etapa, o que facilitaria ainda mais o processo. Já que o treinamento das redes poderia ser realizado em nível de povoamento, com a adição de outras variáveis visando aumentar a precisão das estimativas.

Outra questão que merece destaque, apesar de não ser objeto deste estudo, é a situação em que há a necessidade de calcular o índice de local para ajuste de modelos de crescimento e produção, mas a classificação foi realizada por meio de rede neural artificial de treinamento não supervisionado. Uma alternativa é considerar cada classe obtida como um estrato e aplicar um método tradicional de construção de curvas de índices de local. Outra alternativa é utilizar uma rede neural artificial de treinamento supervisionado para estimativa do índice de local em nível de povoamento e usar as classes obtidas previamente como uma variável categórica. A adição

dessa variável categórica torna a ferramenta adequada para estimativas de índice de local em povoamentos onde a tendência de crescimento em altura dominante seja anamórfica ou polimórfica.

Por fim, na prática, a escolha do método é subjetiva. Em geral, o usuário dará preferência ao método mais prático e ou o que se tenha maior domínio. A comparação dos métodos neste trabalho não teve o objetivo de selecionar um método mais adequado, mas verificar se havia consistência nos resultados obtidos. Era esperado que o método da predição dos parâmetros fosse mais discrepante dos demais, tendo em vista que o padrão de crescimento em altura dominante do povoamento é polimórfico. No entanto, isso não inviabiliza o uso do método da curva-guia, nessa situação, pois as diferentes tendências de crescimento foram consideradas na estratificação para ajuste do modelo utilizado.

6 CONCLUSÕES

O povoamento florestal apresenta padrão polimórfico de crescimento em altura dominante.

O conjunto de variáveis volume total com casca, área basal, altura total, altura dominante, diâmetro quadrático médio e número de árvores por hectare foi a entrada que proporcionou maior porcentagem de acerto na classificação por meio de rede neural artificial, segundo a análise discriminante.

Segundos os critérios de comparação adotados, os métodos da curva-guia e equação das diferenças foram idênticos, enquanto pelo método da predição dos parâmetros as classes obtidas foram diferentes dos demais métodos tendo em vista o padrão polimórfico de crescimento em altura dominante.

A classificação por meio de rede neural artificial foi eficiente quando comparada aos demais métodos em termos de porcentagem de coincidência na alocação dos talhões por classe de capacidade produtiva, área e estoque volumétrico.

REFERÊNCIAS

- ARCOVERDE, G. F. B.; ALMEIDA, C.M.; XIMENES, A.C.; MAEDA, E.E.; ARAÚJO, L.S. Identificação de áreas prioritárias para recuperação florestal com o uso de rede neural de mapas auto-organizáveis. **Bol. Ciênc. Geod. (Online)**, Curitiba, v. 17, n. 3, 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-21702011000300004&lng=en&nrm=iso>. Acesso em: 14/05/2014.
- ASSMANN, E. **The principles of forest yield study: studies in the organic production, structure, increment and yield of forest stands.** Pergamon Press, 1970.
- BEZERRA, A. F. **Modelagem do crescimento e da produção de povoamentos de *Tectona grandis* submetidos a desbaste.** 2009. 60 f. Dissertação (Mestrado) – Universidade Federal de Viçosa, Viçosa, 2009.
- BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G.. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Rev. Árvore**, Viçosa, v. 37, n. 4, p. 639-645, 2013 .
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações.** 2.ed. Rio de Janeiro: LTC, 2007.
- BRAGA, A. P.; CARVALHO, A.P. L.F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações.** 1 ed. Rio de Janeiro: LTC, 2000.
- BRAGA, F. A.; BARROS, N. F.; SOUZA, A. L.; COSTA, L. M. . Características ambientais determinantes da capacidade produtiva de sítios cultivados com eucalipto. **Rev. Bras. Ciênc. Solo**, Viçosa, v. 23, n. 2, p. 291-298, 1999.
- BRUM, E. T. **Relações entre a altura dominante e fatores do sítio, em povoamentos de *Pinus elliottii* Engelm. na região de Ponte Alta do Norte, SC.** 1979. 178 p. Dissertação (Mestrado) - Universidade Federal do Paraná, Curitiba, 1979.
- CALDEIRA, M. V. W.; TONINI, H.; HOPPE, J. M.; WATZLAWICK; L. F.; SELEE, G. L. Definição de sites em povoamentos de *Pinus elliottii* Engelm. na região de Encruzilhada do Sul, RS. **Ciência Florestal**, Santa Maria, v.6, n.1, p.1-13. 1996.
- CAMARGOS, J. L. de. **Modelagem do crescimento e produção florestal com número variável de parcelas mensuradas.** 2017. 118 p. Dissertação (Mestrado) - Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2017.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração Florestal: perguntas e respostas.** 4.ed. Viçosa: UFV, 2013. 605p.

CARRIJO, J. V. N. **Uso de índices morfométricos na classificação das unidades produtivas em povoamentos de *Eucalyptus urophylla* no estado de Goiás.** Trabalho de Conclusão de Curso - Universidade de Brasília, Brasília, 2016.

CARVALHO, A. P.; MENEGOL, O.; OLIVEIRA, E. B.; MACHADO, S. A.; POTTER, R. O.; FASOLO, P. J.; FERREIRA, C. A.; BARTOSZECK, A. Efeitos de características do solo sobre a capacidade produtiva de *Pinus taeda*. **Boletim de Pesquisa Florestal**. Colombo, n. 39, p.51-66, 1999.

CLUTTER, J. L.; FORTSON, J. C.; PIENAAR, L. V.; BRISTER, G. H.; BAILEY, R. L. **Timber management: a quantitative approach**. New York: John Wiley e Sons, 1983. 333 p.

COSENZA, D. N.; LEITE, H. G.; MARCATTI, G. E.; BINOTI, D. H. B.; ALCÂNTARA, A. E. M.; RODE, R. Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. **Scientia Forestalis**. Piracicaba, v. 43, n. 108, p. 955-963, 2015.

COSTA, J.A.F. **Classificação automática e análise de dados por redes neurais auto-organizáveis**. Tese de Doutorado - Faculdade de Engenharia Elétrica e de Computação, Unicamp, 1999.

COSTA, J.A.F; GONÇALVES, M.L; NETTO, M. L.A. Visualização e análise de agrupamentos usando redes auto-organizáveis, segmentação de imagens e índices de validação. **Journal of the Brazilian Neural Network Society**, vol. 9, Iss. 2, p. 91-103, 2011.

COSTA, J.A.F., NETTO, M.L.A. Segmentação do SOM baseada em particionamento de grafos. In: **Proc.VI Brazilian Conf. on neural networks**, São Paulo, p. 451-456, 2003.

CURRY, B.; DAVIES, F.; PHILLIPS, P.; EVANS, M.; MOUTHINO, L. The Kohonen self-organizing map: an application to the study of strategic groups in the UK hotel industry. **Expert Systems**, 2001, 18 (1). p. 19-31.

DENNIS, J. E.; GAY, D. M.; WELSCH, R. E. An adaptive nonlinear least-squares algorithm, **NBER Working Papers Series**, Cambridge, 1977.

DIAS, A. N.; LEITE, H. G.; NOGUEIRA, G. S.; RODRIGUES, F. L. Avaliação de métodos de ajuste de curvas de índice de local em povoamento de eucalipto desbastados. **Revista Árvore**, Viçosa, v. 29, n. 5, p. 741 – 747, 2005.

ELFVING, B.; KIVISTE, A. Construction of site index equations for *Pinus sylvestris* L. using permanent plot data in Sweden. **Forest Ecology and Management**, p. 125-134, 1997.

FARIA, E. L; ALBUQUERQUE, M.P; ALFONSO, J.L.G.; ALBUQUERQUE, M.P.; CAVALCANTE, T.P. **Introdução ao Toolbox de redes neurais de Kohonen**. Centro Brasileiro de Pesquisas Físicas (CBPF)- Notas Técnicas, CBPF-NT-001/10. Rio de Janeiro, 2010.

Disponível em: <http://cbpfindex.cbpf.br/publication_pdfs/nt00110.2010_04_19_16_17_29.pdf>. Acesso em: 20/05/2014.

FIGUEIREDO, E. O. **Avaliação de povoamentos de teca (*Tectona grandis* L.f.) na microrregião do Baixo Rio Acre**. 2005. 301p. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras, 2005.

GALETE, L.B. **Um método para agrupamento em fluxo de dados utilizando algoritmo som**. 2012. 82 f. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Paraná, Curitiba, 2012.

GARCIA, O. **Índices de sitio preliminares para eucalipto**. Instituto Forestal - Proyecto CONICYT - FONDEF 2/33 Chile, 1995. 17 pp. Disponível em: <http://biblioteca.infor.cl/DataFiles/18566.pdf>. Acesso em: 28/02/2017.

GIRARDIN, J. L.; BROQUEN, P.. El crecimiento de *Pinus ponderosa* Dougl. Y *Pseudo tsugamenziesii* (Mirb). Dougl. en diferentes condiciones de sitio. **Bosque**, Valdivia, v.16, n.2, p.57-67, 1995.

GONÇALVES, J. L. de M.; DEMATTÊ, J. L. I.; COUTO, H. T. Z. do. Relações entre a produtividade de sítios florestais de *Eucalyptus grandis* e *Eucalyptus saligna* com as propriedades de alguns solos de textura arenosa e média no estado de São Paulo. **IPEF**, n.43/44, p.24-39, 1990.

GONÇALVES, M.L; NETTO, M.L.A.; ZULLO JR, J.; COSTA, J.A.F. Classificação não-supervisionada de imagens de sensores remotos utilizando redes neurais auto-organizáveis e métodos de agrupamentos hierárquicos. **Revista Brasileira de Cartografia**, nº 60/1 : 17-29, 2008.

GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: Bookman, 2011.

HAYKIN, S. **Redes neurais: princípios e práticas**. 2 ed. Porto Alegre: Bookman, 2001. 900p.

HIGA, R. H.; GIACHETTO, P.F.; YAMAGISHI, M.E.B.; IBELLI, A.M.G.; REGITANO, L.C.A.; CARDOSO, F.F. Análise de agrupamento de dados de expressão gênica na Rede Genômica Animal. **Embrapa Informática Agropecuária**. Comunicado Técnico, 101, Campinas, SP, Dezembro de 2010.

KLOBUCAR D., SUBASIC M. Using self-organizing maps in the visualization and analysis of forest inventory. **iForest Biogeosciences and Forestry (online)**, v. 5, pp. 216-223, 2012. Disponível em: <<http://www.sisef.it/iforest/pdf/?id=ifor0629-005>>. Acesso em: 14/05/2014.

KOHONEN, T. **Self-organizing maps**. 3rd Ed., Berlin: Springer, 2001. 501 p.

LEITE, A. T. O. **Classificação da capacidade produtiva de povoamentos utilizando altura dominante, altura média e diâmetro médio das árvores.** Trabalho de Conclusão de Curso- Universidade Federal de Viçosa (UFV). Viçosa, Minas Gerais, 2014.

LEITE, A.Y.C. M.; MATOS, L.N.; SILVA, M.A.S. Utilização de rede neural som para visualização e análise de dados censitários. In: 58. **Reunião Anual da Sociedade Brasileira para o Progresso da Ciência** – 58. SBPC. Anais ..., 16-21 julho, Florianópolis, SC. Florianópolis: SBPC, 2006.

LEITE, H.; CASTRO, R.; SILVA, A.; JÚNIOR, C.; BINOTI, D.; CASTRO, A. F.; BINOTI, M.. Classificação da Capacidade Produtiva de Povoamentos de Eucalipto Utilizando Diâmetro Dominante. **Silva Lusitana**, Lisboa, v. 19, n. 2, p. 181-195, 2011.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**. n. 4, pp. 18-36, 2009. Disponível em: <<http://www.fsma.edu.br/si/sistemas.html>>. Acesso em: 14/05/2014.

LYNN, S. **Self-organising maps for customer segmentation using R**. 2014. Disponível em: <<http://shanelynn.ie/index.php/self-organising-maps-for-customer-segmentation-using-r/>>. Acesso em: 15/04/2015.

MACHADO, S A. Curvas de índice de sítio para plantações de *Pinus taeda* na região central do Estado do Paraná. **Revista Floresta**, v. 11, n. 2, p. 4-18, 1980.

MARCOLIN, M. **Polimorfismo e tamanho da amostra para construção de curvas de índice de sítio para *Pinus taeda* L., no Segundo Planalto Paranaense.** 1990. 64 f. Dissertação (Mestrado) – Universidade Federal do Paraná, Curitiba, 1990.

MARTINS, B. F. **Uso da tecnologia LiDAR para classificação da capacidade produtiva de povoamentos de eucalipto.** 2015. 24 f. Dissertação (Mestrado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa, 2015.

MENEGOL, O. **Índice de sítio e a relação entre altura dominante e teores nutricionais das acículas em povoamentos de *Pinus elliottii* var *elliottii* no segundo planalto paranaense.** Dissertação (Mestrado) - Universidade Federal do Paraná. Curitiba, 1991. 74p.

NOLETO, L. F. **Métodos de segmentação de mapas auto-organizáveis para análise de agrupamento.** Trabalho de Conclusão de Curso- Universidade Federal de Santa Catarina. Florianópolis, Santa Catarina, 2007.

OLIVEIRA, M.L.R. de; LEITE, H.G.; NOGUEIRA, G.S.; GARCIA, S.L.R.; SOUZA, A.L. de. Classificação da capacidade produtiva de povoamentos não desbastados de clones de eucalipto. **Pesquisa Agropecuária Brasileira**, v.43, p.1559-1567, 2008.

OLIVEIRA, R. A. **Classificação de sítios em plantações de eucalipto pelo método de índices de local e por classe de solo e precipitação.** 1998. 84 p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Viçosa, 1998.

PELLISSARI, A. L.; FILHO, A. F.; MACHADO, Sebastião do Amaral; CALDEIRA, Sidney Fernando. Geostatistical modeling of site index classes in teak stands. **SOP Transactions on Statistics and Analysis**, v. 1, n.2, p. 74-85, 2014.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RStudio Team (2016). RStudio: integrated development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

RABELO, L.D.O. **Modelagem do crescimento e produção florestal com a estratificação de povoamentos empregando o método k-means.** 2014. 79 p. (Dissertação - Mestrado em Ciência Florestal) - Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2014.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. **Revista Ceres**, Viçosa – MG, v. 40, n. 228, p. 176-195, 1993.

RIBEIRO, E. F. **Caracterização de imagens utilizando redes neurais artificiais.** 2009. Dissertação (Mestrado)-Universidade Federal de Uberlândia, Uberlândia, 2009. 156 f.

RIBEIRO, M. S. **Geoestatística utilizada no planejamento estratégico de florestas de produção de *Pinus taeda* L..** 2015. Dissertação (Mestrado)-Universidade Federal do Paraná, Curitiba, 2015. 83 f.

SANTOS, J. S. **Estratificação de florestas de eucalipto com base na forma do fuste das árvores.** 2014. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, 2014. 72 f.

SCAVINSKI, V. **Projeção da produção utilizando curvas de sítio anamórficas e polimórficas em povoamentos de *Eucalyptus grandis* W. Hill. Ex. Maiden.** 2014. Dissertação (Mestrado em Ciências Florestais) - Universidade Estadual do Centro-Oeste, Irati, 2014. 61 f.

SCHNEIDER, P. R; SCHNEIDER, P. S. P. **Introdução ao manejo florestal.** 2 ed. Santa Maria: FACOS-UFSM, 2008. 566 p.

SCHUMACHER, F. X. A new growth curve and its application to timber yield studies. **Journal Forestry**, Bethesda, v. 37, p. 819–820, 1939.

SCOLFORO, J.R.S. **Biometria Florestal**: modelos de crescimento e produção florestal. Lavras: UFLA/FAEPE, 2006. 393 p.

SCOLFORO, J. R. **Curvas de índice de sítio para *Pinus caribaea* var. *hondurensis***. IPEF, n.45, p. 40-47, 1992.

SELLE, G. L.; SCHNEIDER, P. R.; FINGER, C.A. G. Classificação de sítio para *Pinus taeda* L., através da altura dominante, para a região de Cambará do Sul, RS, Brasil. **Ci. Flor.**, Santa Maria, v.4, n.1, p. 77-95, 1994.

SHARMA, M.; AMATEIS, R. L.; BURKHART, H. E. Top height definition and its effect on site index determination in thinned and unthinned loblolly pine plantations. **Forest Ecology and Management**, Blacksburg, v. 168, p. 163-175, 2002.

SILVA, M. A. S. **Mapas auto-organizáveis na análise exploratória de dados geoespaciais multivariados**. 2004. 120 p. Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2004.

SKOVSGAARD, J. P.; VANCLAY, J. K. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. **Forestry**, v. 81, n. 1, 2008.

SUBEDI, N.; SHARMA, M.; PARTON, J. An evaluation of site index models for young black spruce and jack pine plantations in a Changing Climate. **Climate change research report** 15.Ontario: [s. n.], 2009.

TAQUETTI, M. F. **Método da predição de parâmetros em modelos de crescimento em função do índice de local**. Trabalho de Conclusão de Curso - Universidade Federal de Viçosa. Viçosa, 2014.

TONINI, H.; FINGER, C. A. G.; SCHNEIDER, P. R.; SPATHELF, Peter. Crescimento em altura de *Pinus elliottii* Engelm., na região de Piratini no Rio Grande do Sul, Brasil. **Ciência Rural**, Santa Maria , v. 31, n. 3, p. 417-423, 2001.

TONINI, H.; SCHNEIDER, P. R.; FINGER, C. A. G. Crescimento de clones de *Eucalyptus saligna* Smith, na Depressão Central e Serra do Sudeste, Rio Grande do Sul. **Ciência Florestal**, Santa Maria, v.14, n. 2, p. 61-77, 2004.

ULTSCH, A. Knowledge extraction from self-organizing neural networks. In: Opitz, O. ed. **Information and Classification**. Berlin: Springer, 1993. p. 301–306.

VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. SOM Toolbox for Matlab 5. report A57, April 2000. Libella Oy: Finland: SOM Toolbox Team, 45 Helsinki University of Technology, 2000a. 59 p., ill. Disponível em: <<http://www.cis.hut.fi/somtoolbox/package/papers/techrep.pdf>>.

Acesso em: 21/05/2014.

VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. **IEEE Transactions on Neural Networks**, v. 11, n. 3, p. 586–600, 2000b.

VESANTO, J.; SULKAVA, M.; HOLLMEN, J. On the decomposition of the self-organizing map distortion measure. **Proceedings of the Workshop on Self-Organizing Maps**, pp 11-16, Hibikino, Kitakyushu, Japan, 2003.

WEHRENS R.; BUYDENS L.M.C. Self and super-organizing maps in R: the kohonen package. **Journal of Statistical Software**, v. 21, Issue 5, 2007.